

3D打印

三维智能数字化创造

创客实践

中国智造

第三次工业革命

(第2版)

吴怀宇 / 著

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

《经济学人》等主流媒体称 3D 打印将引发“第三次工业革命”。本书从产业经济的宏观视角对 3D 打印、3D 智能数字化、创客、中国智造、全球第三次工业革命这五者的关系进行了详尽讨论。

本书从专业技术的角度对 3D 打印的原理、结构和工艺方法做了详细介绍,包括 10 多种典型成型工艺的优劣分析和比较,手把手、从无到有地组装一台 3D 打印机等。3D 智能数字化是 3D 打印的基础和关键,涉及 3D 计算机图形学、计算机视觉、模式识别、机器学习等领域。本书以通俗易懂、娓娓道来的方式对它们进行了详细讲解。

本书是一本以作者原创观点为指导,融汇众多最新思想,详细讲解 3D 打印和 3D 智能数字化技术原理方法,手把手实战型教学的综合类技术书籍,对每一个操作步骤都进行了图文并茂的详细描述,包括实际运作一家 3D 照相馆的所有技术细节。本书无论对于国内、国外的广大普通用户及技术爱好者,还是高等院校大学生及研究生、学术界、工业界、政府产业经济决策层,都具有重要的参考价值。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

3D 打印:三维智能数字化创造 / 吴怀宇著. —2 版. —北京:电子工业出版社, 2015.1
ISBN 978-7-121-24641-8

I. ①3… II. ①吴… III. ①立体印刷—印刷术 IV. ①TS853

中国版本图书馆 CIP 数据核字(2014)第 248386 号

责任编辑:付 睿

印 刷:中国电影出版社印刷厂

装 订:河北省三河市路通装订厂

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:787×1092 1/16 印张:27.5 字数:720 千字

版 次:2014 年 1 月第 1 版

2015 年 1 月第 2 版

印 次:2015 年 1 月第 1 次印刷

印 数:3500 册 定价:105.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

推荐序一

今年（2013年）4月看过吴怀宇博士给《中国科学报》撰写的关于3D打印的技术评论专栏，感觉写得十分顺畅，现在很高兴得知他的《3D打印：三维智能数字化创造》一书也即将出版了，更是神速！

3D打印近年来在国内外受到广泛关注，网络、报刊、媒体上都有铺天盖地的密集报道。新闻报道虽然可以及时反映3D打印技术的最新进展，但呈现给普通读者的观感难免有些碎片化，不便于获得系统化的了解和认识，尤其是3D打印机几乎“无所不能”地被应用到各个领域，从航空航天、汽车、医疗，到时装、食品、武器等，给人一种眼花缭乱的感觉。《3D打印：三维智能数字化创造》出现得非常及时，是一本很值得读的书。该书不仅对3D打印的来龙去脉、原理结构、各种工艺方法做了系统的阐述，而且还对3D智能数字化的理论方法、应用技术以及最新进展也做了详尽的介绍。将这两个密切相关的新兴学科关联起来进行表述，使得读者能够从整体上把握3D数字化打印的脉络框架。此外难能可贵的是，吴怀宇博士非常耐心地以通俗易懂的文字解释书中出现的每一个术语和技术细节，辅以图文并茂的形式，使得本书将能够被广大普通读者所接受。

近些年来我也非常关注3D打印和社会制造方面的研究进展，并曾于2012年在《中国科学院院刊》上发表过一篇题为“从社会计算到社会制造：一场即将来临的产业革命”的文章。借怀宇的新著，我对3D打印和社会制造谈一点自己的体会和看法，抛砖引玉，同大家一起探讨。

具体说来，源自快速成型和快速制造，以3D打印技术为核心手段的加式制造（Additive Manufacturing），被许多人认为是一项将要改变世界的“破坏性”新技术，已引起全球性的关注。加式制造是相对于减式制造而言的，两者过去都不是严格意义下的制造专业术语。所谓减式制造，即通过模具、车铣等机械加工技术与工具将原材料转化成产品的工艺过程与设备的总称，其特征为利用缩削、减少材料来生产部件。而近十年来，随着快速成型、快速制造、3D打印等技术的成熟与普及，加式制造已成为日益风行的制造专业术语。与减式制造相反，加式制造的主要特征就是利用逐层增加材料的方式生产各种产品，无须模具，因此也被称为无形制造技术（Freeform Fabrication，简称FF或FFF）。

2012年3月，英国《经济学人》杂志以“第三次工业革命”为主题，声称3D打印技术即将引发新一轮的“工业革命”浪潮，并认为生产制造将从大型、复杂、昂贵的传统工业过程中分离出来，凡是能接上电源的任何计算机都能够成为灵巧的生产工厂；人类将以新的方式合作进行生产制造，制造过程与管理模式将发生深刻变革，目前的制造格局必将被打破。

然而，正如蒸汽机促成第一次工业革命是通过引发人类理念的变革来达成的，3D打印机要催生新的产业革命，也必须通过诱发新的人类理念转化来实现。问题是：新的理念是什么？

我们认为，这一新的理念即便不直接是社会制造，也一定与社会制造直接相关。社会制造可使传统企业转变为能够主动感知并且响应用户大规模定制需求的智能企业，其核心就是主动、实时地将社会需求与社会制造能力有机地衔接起来，从而有效地实现需求和供应之间的相互转化。为此，我们必须把社会搜索、社会计算、社会制造等相关的新兴领域有机地结合起来，将互联网、物联网和物流网与3D打印机组成的社会制造网无缝地连接，通过众包等方式使社会民众充分参与产品各个环节的全生命制造过程，促成个性化、实时化、经济化的生产和消费模式，形成新的产业革命。

正如Google依靠大规模的计算机服务器阵列满足人们信息搜索的需求，从而改变人类生活与工作方式一样，我们可以设想未来的3D打印机也将组成大规模的社会制造阵列，实时方便地满足人类

对各种个性化产品的物质需求，使生活和产业中的“长尾效应”常态化，进而更加深刻地改变我们生活的社会。这就是为什么 3D 打印将改变我们的世界，这就是为什么社会制造将带来一场产业革命的真正原因。

中国作为今日世界制造业大国之地位正面临着严峻的挑战，西方媒体甚至直白地宣称：“天将变了”，“未来的制造业将再次回流到先进发达国家”，“美国制造，出口中国”的新时代即将来临！就连美国总统奥巴马也在 2013 年 11 月的国情咨文讲话里特别地强调 3D 打印技术，全美国上下要将其视为拯救美国制造业的希望之光。

参照信息行业的发展历程，我们认为，快速成型相当于 20 世纪 60 年代的专用和大型计算机，3D 打印机则相当于 20 世纪 70 年代的个人 PC 和苹果台式计算机。令人担忧的是，我们在这一新兴领域目前所处的地位，差不多就是半个世纪前我国在世界信息行业所处的地位！

显然，我们必须尽快补上 3D 打印这一课，但我们切不可忘记信息行业在个人计算机出现之后浪潮般的发展进程：Microsoft 的快速崛起，还有随之而来的 Oracle、Yahoo、Amazon、eBay、Google、Facebook、Twitter，国内的百度、阿里巴巴、QQ 和微博等。以目前的情况判断，3D 打印的核心价值将完整体现于社会制造的发展与成熟的过程当中。社会制造对于制造行业而言，就是信息行业中从 Microsoft 至 Amazon 再到 Google 和 Twitter 的一体化合成，可视为虚拟网络世界与真实物理世界的首次完美结合。因此，在关于 3D 打印机之大量媒体渲染的背后，社会制造才应当是我们关注的要点，否则，我们可能错失良机，一误再误，代价将难以估量。

在社会制造的环境中，大批 3D 打印机形成制造网络，并与互联网、物联网和物流网无缝连接，形成复杂的社会制造网络系统，实时地满足人们的各种需求。消费者与企业通过网络世界能够随时随地参与到生产流程之中，社会需求与社会生产能力将实时有效地结合在一起，“想法到产品（Mind to Product）”，“需求就是搜索，搜索就是制造，制造就是消费”将成为现实。因此，社会制造必将极大地刺激社会需求，同时有效地提升整个社会的参与程度，其直接结果就是社会就业率的大幅提高。而且，加速发展社会制造产业，不但能够解除我国长期在模具和材料工业落后受制于人的不利局面，还可以使我国蓬勃发展的社会媒体和网络文化得到进一步的升华，使其成为促进社会经济科学发展的有力工具：从被动到主动，从消极到积极。

另一方面，社会计算也将发挥关键性的作用，从专注社会舆情分析到满足社会经济需求，为社会制造的发展与成功提供有力保障。首先，社会计算为社会制造提供了主动及时地掌握社会需求的必要手段，从而能够在大数据时代环境下直接用数据考察研究各类问题。其次，社会制造涉及人的行为与需求，对许多问题由于时间、经济、法律和道德上的原因无法进行传统的实验，而社会计算则能够以计算实验的方式弥补这一缺陷。最后，社会计算的平行管理与控制为落实社会制造的运营和支持各种决策提供了一个有效的操作平台。

一言蔽之，社会制造的关键就是通过社会计算，主动、实时地将社会需求与社会制造能力有机地衔接起来，从而有效地实现需求和供应之间的相互转化过程。由于存在于社交媒体上的大数据具有动态性、多样性、虚实交互性、复杂性和不确定性等特点，如何计算获得有用的信息，并从中挖掘出一般规律是一个极其具有挑战性的问题。我们必须以物联网、云计算的手段，采用机器学习、数据挖掘、模式识别、人工智能等领域的理论、技术和方法，研发可计算的智能社交媒体数据信息处理机制。为此，必须把社会计算和社会制造这两个密切相关的新兴领域有机地结合起来，这将对提高我国制造业的竞争力、加速产业升级和转型、扩大社会内需、繁荣国家经济，具有至关重要的战略意义。

2007 年，在参加编写《中国至 2050 年先进制造科技发展路线图》的过程中，加式制造引起了我们的注意，但当时由于专家意见不一致，特别是人力、物力和时间的缺乏，只能将加式制造作为实验

室的一个跟踪课题予以关注。然而，3D 打印技术和社会制造的发展速度却大大超过我们的预期。很明显，社会制造是计算机和互联网引发的信息革命之后的又一场产业革命，而且是一场虚实结合的革命，其规模和速度都将是前所未有的，意义重大，并更具挑战性。这场革命对从业人员的素质与专业水平以及运营环境的要求都与我们现行的教育科研和产业管理体制有明显的冲突。如不认真应对，轻则可能发生西方国家所期望的制造业从中国等发展中国家向发达国家回流的现象，重则严重影响中华民族复兴的伟业。

希望像蒸汽机一样，3D 打印机能够通过社会制造的理念和实践，使人类社会再一次从以开发“地下”资源为主的“工业”社会，一步跃入以开发“地上”数据和智力资源为特征的“智业”社会，充分发挥人类共有的智力，使数据真正地成为驱动和支撑大数据时代社会发展的“石油”和“黄金”矿藏。

正如怀宇博士在书中提到的，3D 打印是个技术密集型的行业，需要依托包括信息技术、精密机械和材料科学等多个学科领域的共同发展，才能加速“中国制造”转型升级为“中国智造”的过程，并以此推动“全球第三次工业革命”。在此衷心期望各界人士能够齐心协力、携手并进，共同抓住这次伟大技术变革的历史机遇，一起实现我国科技的跨越式发展。

王飞跃 研究员
中国科学院自动化研究所
复杂系统管理与控制国家重点实验室 主任
IEEE Transactions on ITS 主编、中国自动化学会秘书长

推荐序二

3D 打印火了。

实际上，作为技术本身，3D 打印并不新。30 年前，3D 打印就已出现在大公司和科研院所的实验室里，业界称之为快速成型或增材制造。

成名似乎是在一夜间。2012 年，英国著名经济学杂志《经济学人》发表封面文章，声称 3D 打印将引发全球第三次工业革命。2013 年，“3D 打印”随即成为各大媒体的“宠儿”。

随着探讨的深入，亦有一些不同的声音：如果它真是一项突破性技术，为什么直到 30 年后的今天才开始引发第三次工业革命呢？

从一位公众的角度，我也是带着这个疑问，对全书进行了阅读，希望能从中找到一些答案或线索。

身为这一领域的专业人士，吴怀宇博士用文字还原了真实的 3D 打印及其未来愿景。所谓“第三次工业革命”，其实是“一盘很大的棋”，并不能仅靠一项技术来支撑，而是需要信息技术、先进制造技术、新能源技术、新材料技术、生物技术等众多领域共同发力。当 3D 打印以这些新兴技术为基础，经过 30 年的历练，形成了一个新兴的交叉学科和丰富的技术集群，其对社会产生的影响将足以引发一场综合性变革。

在作者看来，第三次工业革命是以“智能数字化制造及新型材料应用”为代表的崭新时代，其典型特征概括成一个词那就是“智能数字化”。智能数字化技术提高了设计制造工艺的精度和效率；随着数字化车间乃至数字化工厂的出现，生产系统将向着具有感知、决策、执行能力的智能化系统发展。

在这一过程中，制造业原有的发展模式将被改变。3D 打印、智能数字化、新材料以及机器人技术的发展，将使得制造业依靠较少的自然资源和人力资源投入，也能取得良好的经济效益。其中，3D 智能数字化利用计算机来智能地生成数字化的 3D 模型，使得低成本的大规模个性化定制成为可能。3D 智能数字化与 3D 打印技术的完美结合，让设计师和工程师从产品制造工艺的束缚中解放出来，更加专注于产品本身的智力创造，即所谓“想法到产品（Mind to Product）”及“所想即所得”的全新智造时代。

对我而言，最感兴趣的是这本书从产业经济的宏观视角和技术方法的微观视角对 3D 打印、3D 智能数字化、创客、中国智造、全球第三次工业革命这五者的关系进行了翔实的讨论，并以此为逻辑主线将各章节贯穿在一起。

而谁又是 3D 打印直接的推手呢？本书给出了回答——创客（Maker）。

这个名词在国内目前并不为大家所熟知，但实际上，创客运动在欧美已如火如荼。美国《连线》杂志前主编克里斯·安德森在《创客：新工业革命》一书中已做了详细介绍。创客指喜欢动手制作，努力把各种创意转变为现实产品的人。他们会使用 3D 打印机、数控机器、电子电路、激光切割机、3D 智能数字化技术等功能强大的数字桌面工具进行创造。创客，既是工具的发明者，也是工具的使用者。

2008 年，英国一名叫 Adrian Bowyer 的创客发布了第一款开源的桌面级 3D 打印机 RepRap，并把机械设计图纸、电路图纸、控制源代码等无偿放到了网上供人免费下载。这使得原本动辄几十万元的 3D 打印机降价到现在几千元即可买到，从此走入了普通用户家庭。如果没有创客，没有他们的开

源共享精神，现在的 3D 打印机还会这么便宜吗？为此，将创客运动称作第三次工业革命的启蒙运动非常贴切。

公众需要这样一本书，能够将看似高深的 3D 打印深入浅出，又不失专业性地展示出来。怀宇的这本新书语言易懂、图文并茂，根据读者群的不同诉求，划分出了五大派别：操作实战派、技术方法派、商业运作派、大局宏观派、学院理论派，从而增加了这本书的可读性和趣味性。

正如克里斯·安德森所说：“所有重要的科技都是在短时间内被过度炒热，其功能性也被高估；但从长期来看，它们造成的影响却远被低估。”或许不久，“上千元买台 3D 打印机”能被更多的人所接受；也或许，“根据自己的个性化需求打印”遍地开花尚需时日。

但毋庸置疑的是，第三次工业革命已“山雨欲来”，那么中国将在其中扮演一个什么角色呢？面对这一次历史新机遇，中国会依旧只是个追随者吗？这是这本书所引发的更重要的思考之所在。

黄明明
《中国科学报》技术经济周刊 主编

推荐序三

吴怀宇博士请我给他这本关于 3D 打印的书写个序，可能是因为我做了 25 年多的三维计算机视觉研究，而三维视觉和 3D 打印有紧密联系之故。刚好我对 3D 打印很有兴趣，苦于没有系统跟踪，就欣然答应。我利用圣诞和元旦假期，对本书原稿从头到尾读了一遍。吴博士用笔风趣幽默，对三维智能数字化创造充满激情，对 3D 打印看似庞大的生态系统提纲挈领，对深奥的理论和细节能深入浅出，对这项新工业的经济技术发展的现状和趋势娓娓道来、引人入胜。他对 3D 打印、三维智能数字化、创客、中国智造、全球第三次工业革命之间的内在紧密关联有着深入的研究。我从读前言开始就被吸引，爱不释手，在此强烈推荐。你们不会失望的。

3D 打印早已从理论变为现实。3D 打印机近几年变得越来越便宜，1000 美元就能买到一台三维桌面打印机。微软推出的 Windows 8.1 操作系统也开始支持 3D 打印，其操作简单，能和打印文本一样轻松实现 3D 打印。另一方面，除了日常物品，在高端商业物品和部件制造领域，比如燃气轮机，3D 打印技术也都在发生日新月异的进展，而且很有可能在不远的将来为了医疗目的对人类器官进行生物打印。

尽管 3D 打印技术有了长足发展，但仍有很多需要改进和完善的地方，期待着我们去创新，去发现商机。我们可从 3 个层次考虑：打印技术、内容创造和生态系统。

- 打印技术：从设备角度来看，目前，3D 打印是一个添加式制造过程，而传统制造是一个削减式生产过程，能否将两者整合？在材料上，目前还基本停留在“黑白”，什么时候能打印出栩栩如生、栩栩如生的物件？从软件支持方面，能否在保质的基础上自动编辑打印内容以减少耗材？
- 内容创造：没有丰富的、吸引人的、有用的 3D 数字化产品，3D 打印不可能产生巨大的消费市场，也就不可能推动全球第三次工业革命。要想使产品设计大众化，离不开一个简单易用的三维数字化设计软件。能否用类似 Kinect 的传感器，开发一种自然用户界面（NUI），让用户像创作雕塑一样在数字空间里自然直观地设计产品？能否轻松地将现实世界的物体转换成数字模型？如何在现实物体上增加虚拟成分？这些都有赖于计算机图形学、计算机视觉、模式识别、机器学习等交叉学科的知识，而本书也通俗易懂地介绍了这些技术。
- 生态系统：一个新兴行业的生态系统需要精心呵护，既不能扼杀创客们的热情和创造性，也不能完全放任自由。随着越来越多的创新产品和个人信息数字化及商业行为云端化，如何保护知识产权？如何保护信息安全？如何保护隐私？……这里既需要政策的支持，也需要研究和创新，还有很多的商业机遇。

本书讲 3D 打印，不光对 3D 打印机的原理和结构有一个系统的描述，更强调三维智能数字化技术对推动全球第三次工业革命的核心作用。只要我们抓住机遇，转换思维，从传统制造的束缚中解放出来，学习、研究并牢牢掌握三维智能数字化的核心技术，发挥我们的想象力和创造力，在多学科碰撞中产生灵感和火花，我们一定能超越“中国制造”的成就，创造出更大的“中国智造”的辉煌。

张正友

微软研究院首席研究员，研究经理

国际电气电子工程师学会院士（IEEE Fellow）

美国计算机协会院士（ACM Fellow）

推荐序四

3D printing is an emerging technology that has attracted much attention in recent years. The profound applications of this technology are likely to revolutionize many aspects of our life, ranging from rapid manufacturing, mass customization to health care devices. As discussed and demonstrated by the author, it is likely to make the third wave of industrial revolution.

The timely topics covered in this book provide an excellent introduction of topics related to 3D printing and the involved techniques. The author first gives a thorough review of the latest development of 3D printing and draw connections to the future of manufacturing in China. All the important techniques in 3D printing are then discussed and explained in the following chapters. In particular, the author uses 3D photography as an example to discuss its relevance to 3D printing. In addition, the author explains several computer vision algorithms on how to recover 3D model from images. The future of 3D printing is then discussed with numerous examples to motivate research problems for readers to ponder. For completeness, the author also discusses the underlying mathematics and algorithms related to 3D reconstruction.

In summary, this book provides an excellent introduction of the emerging and important topics of 3D printing. The discussed material not only covers topics of great interest to all readers but also algorithms for practitioners in computer vision. I highly recommend this book to readers interested in 3D printing or computer vision.

Ming-Hsuan Yang

Associate Professor

Electrical Engineering and Computer Science

University of California at Merced, USA

杨明玄

美国加州大学 终身教授

IEEE Transactions on PAMI、IJCV 副主编，ICCV/CVPR/AAAI 领域主席

推荐序五

3D Printing, proclaimed by some as a symbol for the third industrial revolution, has been a very hot topic in recent years. While there are literally over one thousand books about 3D printing that are currently available for sale on Amazon.com, what sets this book apart, I believe, is its comprehensiveness. It covers a broad range of topics from 3D printing principles, hardware, digitization/modeling, to practices. In particular, it emphasizes the software aspect of 3D printing, which is 3D digitalization and modeling. There is no 3D printer that can produce a 3D object without a digital 3D model. This software, i.e., modeling, aspect is somehow overlooked by many 3D printing literatures. I am very happy to see a book that gives a comprehensive and balance treatment of all aspects involved in the 3D printing pipeline.

This book is well suited for both beginners and intermediate practitioners. Beginners can learn the basic principles of 3D printing involving both hardware and software. For intermediate practitioners, who already know the basics, this book provides an under-the-hood view of 3D printing hardware and modeling software. In this regard, this book is particularly useful for innovators who want to develop the next kill-app for 3D printing. The various modeling tools and optimization algorithms introduced in this book are the building blocks for 3D printing applications, such as 3D photography, which is used as the driving example.

The author of this book, Dr. Wu, is an expert in 3D modeling and its applications to 3D printing. The writing style is easy to follow, with many examples and references to related literatures. It is a joy for me to read this book. I highly recommend this book for anyone who is interested in learning 3D printing and applications.

Ruigang Yang

Professor, University of Kentucky

3D 打印被称为是第三次工业革命的象征。它一直是近年来非常热门的话题。目前在 Amazon.com 有很多有关 3D 打印的书籍。本书的最大特点,我认为,是它的全面性。它涵盖了 3D 打印的全部过程,从打印原理、硬件、数字化造型,到以 3D 照相为例子的实践,特别的是它对三维数字化建模软件的强调。三维数字模型是任何 3D 打印的源泉,没有数字模型的 3D 打印只能是缘木求鱼。3D 打印硬件对建模软件的依赖性是被市面上很多 3D 打印书籍或多或少忽视的一个问题,我很高兴地看到本书对 3D 打印的整个流程给出了一个全面、平衡的阐述。

本书非常适合初学者和中级从业人员。初学者可以全面地学习 3D 打印的基本原理,其中包括硬件的和软件的。对于已了解基本知识的中级从业者,本书提供了对 3D 打印硬件和数字化建模软件的深入分析。在这方面,本书对开发应用程序的 3D 打印创新者特别有用。另外,本书对各种建模工具和优化算法进行了详细的介绍和分析,这些是开发新的 3D 打印应用的必备知识。

本书的作者,吴博士,是在三维数字化建模及其在 3D 打印上应用的专家。他的文风深入浅出、通俗易懂。书中有很多示例并详细列出了大量相关参考文献。我强烈地向任何有兴趣学习 3D 打印及其应用的人士推荐本书。

杨睿刚

美国肯塔基大学 终身教授

推荐序六

计算机发展到今天，其最成功的应用集中在信息的处理、储存和传递。这些应用极大地改变了人们的生活。然而，随着计算机功能的逐步增强和计算技术的进一步发展，越来越多的应用已经不局限于从信息到信息，而是将信息用于改变客观世界以及人与客观世界的交互。3D 打印就是这样的技术。这样的技术无疑将会对我们的生活产生更大的影响。

3D 打印一反传统的切削去除材料加工方法，采用增材制造技术，可以加工出任意复杂形状的物体。这项技术经过 30 年的发展，从材料的性能到加工的精度都有了很大的改进。另一方面，人工智能的研究经过几十年的波折，终于有了新的突破，给 3D 打印提供了强有力的软件支持。正是这两方面的发展汇聚在一起才使得人们相信一个全新的设计和制造方式正在形成，并将彻底地改变现有的工业结构。

要想弄懂 3D 打印的过程并有效地解决实际问题需要掌握硬件和软件两方面的知识，而软件方面所涉及的范围则尤其广泛。吴怀宇博士的这本书的独到之处是综合了这两方面的知识并且用通俗易懂的语言解释出来。这正是当前大多数想要了解这个新兴领域的读者所需要的。作者抓住了“三维智能数字化”这个关键的要素，详细地介绍了基于当前最新研究成果所提供的工具。书中既有概念和算法的描述，又有编程和软件的指导，还有关于具体操作的问题解答。

三维智能数字化的核心目标是客观世界的物体建立计算模型。这样的模型为 3D 打印提供了理想的工具。近年来这个处于计算机视觉、计算机图形学和模式识别交叉处的学科取得了令人兴奋的成果，也吸引了一大批有才华的研究人员。吴怀宇博士在这个领域做了广泛的工作并取得了出色的成绩。本书体现了他对这个学科的深入理解。

如果你想了解 3D 打印并自己动手制作模型，或者你想进一步学习 3D 算法和相关的数学方法，就请阅读本书吧！

Shu Chang (舒畅)
Senior Research Scientist (资深科学家)
National Research Council of Canada (加拿大国家研究委员会)

前言

《诗经·小雅·鹤鸣》有云：“它山之石，可以攻玉”。本书取名《3D 打印：三维智能数字化创造》，灵感源自于笔者发表在《光明日报》上的一篇 3D 打印综述文章的标题《3D 打印：智能数字化》。其实笔者最初提交的是一个又长又绕口的题目，非常感谢《光明日报》的编辑以金钢钻般的犀利进行了打磨，使 3D 打印的本质顿时“彰明较著”，也让我那篇文章的条理和纹路立刻清晰了许多！所以本书也受此启发，起了一个相似的名字。

相信大多数读者在拿起这本书的时候，对 3D 打印或多或少已有耳闻。确实，自从《经济学人》、《福布斯》、《纽约时报》等欧美主流媒体声称 3D 打印将引发第三次工业革命开始，全世界各类媒体都对 3D 打印做了大量的跟踪报道。那么，为什么欧美这么看好 3D 打印？3D 打印又为什么会引发一场新工业革命？3D 打印不是 30 年前就有了吗？那时只是一种快速成型工具而已，难道一种“新瓶装旧酒”的工具就会引发一场全球范围内的工业变革？这场变革与我们中国的制造业会有关联吗？此外，媒体上经常报道国外“创客”通过 3D 打印机造出了创意新奇的作品，可当我们也深受鼓舞买回一台，却会发现 3D 打印远没有 2D 打印轻松，其中最头疼的事情莫过于要设计和处理所谓的 3D 数字化模型了。那么，3D 数字化和 3D 打印到底是什么关系？我们又该如何轻松应对呢？

以上这些最基本的问题，实际上已经引出了多个主题，归纳一下：有 **3D 打印**、**3D 智能数字化**、**创客**、**中国智造**、**全球第三次工业革命** 这 5 个关键词，而且都互有关联。忽略掉其中任何一个，都无法完整地回答读者的上述诸多问题。这是摆在本书面前的一个艰巨的任务。倘若选择性忽略，只讨论其中的某一个或几个方面，则 3D 打印机与一般人眼中的 2D 打印机或一台普通机床又有什么差别呢？3D 打印无须模具就可加工任意复杂的中空形状，用户也无须掌握各种复杂的制造工艺和加工技能，这样大幅降低了制造业的技术门槛。3D 打印的巨大威力虽然源于技术，但其产生的重要影响力却又远超于此。

依我看来，欧美现在之所以看好 3D 打印，主要是希望将制造业回流到欧美，而不是继续转移到中国和印度。2007 年爆发的全球金融危机，根源在于美国重视房地产、金融、消费等第三产业的发展而将大量的制造业外包给了其他国家，导致自身产业空心化问题日益严重。3D 打印这种快速成型制造技术最近几年的突然火爆，有一个重要原因和转折点，那就是 2008 年**创客**们发布了第一款完全开源的个人 **3D 打印机** RepRap，并把机械设计图纸、电路图纸、智能控制代码无偿放到了网上供人免费下载。几年下来，原本极其昂贵（几十万元起价）的 3D 打印机降到现在几千元即可买到，变得大众化，由此掀起了“个人智造”、“家庭智造”、“网络社区智造”的热潮。欧美正是希望借创客运动和“全民智造”的东风，激发国民的创造精神，上下齐心来实现这次战略大转移。2014 年 1 月，3D 打印的激光金属烧结技术也将因专利到期而开源，这将为 3D 打印的发展注入更大活力。

与此同时，我国政府也非常渴望将原本处于产业链低端的“中国制造”转型为“**中国智造**”，从加工组装环节升级到上游的设计研发环节。“中国智造”的核心在于智能化和数字化（简称“智

能数字化”），不仅要建立数字化工厂提高各种设计制造工艺的精度和效率，同时要使生产系统向着具有感知、决策、执行能力的智能化系统发展，以做大做强“高端制造”。“中国智造”在**3D打印**产业上的竞争力可以通过发展**3D智能数字化**来提升。实际上，“当今世界是平的”，在经济全球化的背景下，中国制造业的深度发展离不开全球市场化布局。因此，“中国制造”向“中国智造”转型升级历程，实际上也是共同推动和实现“**全球第三次工业革命**”的过程，并将在其中扮演越来越重要的角色。

具体来说，第三次工业革命是以智能数字化制造及新型材料应用为代表的的一个崭新的时代，具体特点可描述为：智能数字化、分布式网络化、个性定制化、绿色可持续化，典型特征为“智能数字化”。3D打印、智能数字化、新材料以及机器人技术的发展，将极大地改变制造业原有的投入模式，使得依靠较少的自然资源和人力资源投入，就能取得良好的经济效益，并将远离产品千篇一律的大规模制造模式，向更具个性化的定制规模发展。

以上就是全书的基本思路和逻辑线索。下面，具体介绍一下本书的主要内容。

本书首先从产业经济的宏观视角对3D打印的发展现状和未来进行了详尽的讨论。为了能使读者对3D打印、3D智能数字化、创客、中国智造、全球第三次工业革命之间的内在紧密关联有比较深入的理解，我们对这五者的相互作用和关系进行剖析。

3D打印将虚拟的智能数字化技术与实实在在的工业产品桥接在一起，跨越了虚拟的比特世界和实体的原子世界之间的鸿沟。为了让读者对3D打印有透彻的了解，我们从专业技术的角度对3D打印的原理结构、成型工艺和实际操作进行了详细介绍，包括对10多种典型的成型工艺进行优劣分析和比较，乃至手把手地、从无到有地组装一台3D打印机，以便让大家看得清清楚楚、明明白白。本书对每一个操作步骤都进行了图文并茂的详细描述，包括实际运作一家3D照相馆的所有技术细节。

3D智能数字化是3D打印的“孪生兄弟”，通过利用计算机来智能化地设计或获取一个3D数字化模型，以便输出到3D打印机。这是本书要讨论的重点所在：为了让用户“所想即所得”地进行数字化创造，计算机需要知道如何更好地生成形状，即能够智能地理解用户的意图。

我们可以使用智能数字化设计软件，从无到有地设计3D数字化产品。最普通的方法是采用传统的建模工具进行实体建模和曲面建模。而手工建模是一件比较烦琐、费时的工作，研究人员于是推出了参数化建模、直接建模工具来减轻设计负担。更加智能化的是编程式设计，计算机把形状的设计过程描述成一系列有特定顺序的操作步骤，有点像按照食谱而不是最终的外观来制作蛋糕。编程式智能设计可以轻易地在这个蛋糕上绘制几百万个规则的精美图案，而这对于手工设计来说犹如噩梦。

为了生成更加丰富多变的个性图案，还可采用复杂的生长式智能系统，即所谓的过程建模。智能化达到一定层次后，更可以让设计的形状根据未知环境实时调整，适应各种物理和美学约束条件。比如，基于算法的智能设计软件能够根据物理环境（如在月球上）调整建筑结构的空间形状，以此来动态获得一个最优的设计形状，从而使建筑结构更加稳定。

当然，并非人人都有能力自己设计3D形状，因此3D智能数字化的另外一种方法就是3D扫描（俗称3D照相），基于计算机视觉、计算机图形学、模式识别与智能系统、光机电一体化

控制等技术对现实存在的 3D 物体进行扫描采集，以获得逼真的数字化重建。在获得数字化模型之后，通常还需要进行个性化编辑定制。特别是对于“大批量定制”，如为一万名用户打印定制个性化的眼镜、服装、帽子、鞋子，则需应用智能化数字技术，如采用视觉计算方法，利用摄像头自动采集、分析提取每位用户的体貌个性特征，进行匹配和定位，并自动根据视觉美感进行形状设计、颜色肤色搭配等，可极大地缩减定制周期。

以开办一家 3D 照相馆为例，这是 3D 智能数字化的典型案例。首先需要对人体进行 3D 扫描或根据多视角照片进行立体重建，然后利用数字几何处理的方法对缺失和噪声数据进行修补，并拼接得到一个完整的 3D 模型。其中头发的快速修复就是一个值得研究的课题，涉及视觉计算技术。此外，用户很可能还希望对 3D 人体形状或表情进行美化、编辑、修改、迁移等，这涉及图形图像、模式识别、机器学习等多个领域。在输出打印前，还涉及形状的自平衡处理、形状分析以提高表现力、大尺寸形状的自动分块、形状优化生成轻质结构以节省耗材、利用增强现实预览融入环境的效果等。

当 3D 数字化模型变得跟目前的 MP3 歌曲一样普及甚至泛滥时，又会遇到如何快速检索的难题。不像 MP3 那样可以通过歌名与歌手名这些结构化的文本信息来定位，3D 模型的检索要复杂得多，涉及非结构化数据的特征提取、相似度度量以及分类算法的设计。更让人头疼的是，在这个大数据时代，我们将被信息的海洋淹没而变得迷失，以至于都不知道每天应该挑选哪些 3D 模型打印出来。这时，通过对大数据的挖掘，个性化推荐系统可以对你的个性偏好进行分析，把你可能会感兴趣的 3D 模型推荐给你。其中，深度学习这种模拟人类大脑进行智能分析学习的方法，将获得越来越广泛的应用。

通过智能感知设备，3D 打印机还可控制制造的行为，对打印的过程进行实时监控，然后根据反馈信息随时做出调整。也就是说，这台 3D 打印机具有学习和控制的能力。将来，通过把人工智能从计算机拓展到现实世界，还可打印具备感知和学习能力的智能物品。此时，3D 打印机就是新一代智能机器人，它们能设计、制造、修理、回收其他机器，甚至能够改进和升级机器自身，达到“机器制造机器”的新境界。

可以说，3D 智能数字化技术是 3D 打印实现“规模定制”的基础和关键所在。因此，本书详细讨论了上面提到的各种 3D 智能数字化理论及其实现方法（如 MVS、SVM、AAM、AdaBoost、粒子滤波、Mean Shift、Visual Hull、深度学习），涉及 3D 计算机图形学、计算机视觉、模式识别、机器学习。我们面向 3D 打印和 3D 数字化行业人士，将这些非常专业化的智能算法理论以通俗易懂的方式娓娓道来，这也是本书的一大特色。

“创客”不仅创造了个人 3D 打印机，同时也是第三次工业革命的启蒙者。这是任何一本 3D 打印书籍都绕不开的话题，因此，我们详细介绍了创客，并专门开设一章介绍四轴飞行器的 DIY 制作，以实例的方式讲解创客们喜爱做的东西，将 3D 打印、智能数字化技术这些先进的工具融入到创客实践当中。

综上，本书是一本以最新视角阐述新工业经济发展趋势、详细讲解 3D 打印与 3D 智能数字化技术原理方法、手把手实战型教学的综合类技术书籍，因此无论对于国内还是国外的广大 3D 打印爱好者、学术圈、工业界、政府产业经济决策层均具有重要的参考价值。

另外，本书提供丰富的网络资源下载，其中的内容包括：Ultimaker 原理图纸、3D 模型头发修复视频教程、四轴飞行器完整资料。如有需要，读者可在 www.broadview.com.cn/22063（博文视点官网）和 <http://www.sigvc.org/why/book/>（作者主页）下载。

凭一己之力是无法完成本书的，在此要衷心感谢多年来一直关心和支持我的师长、朋友、同事和学生。本书在写作过程中得到了汪凌峰博士、王颖博士、刘利刚教授、吴毅红研究员、邓小明副研究员、汪国平教授、唐俊副教授、王俊、王润元、张华、隋伟、赵松、沙金正、李成华、吴挺的帮助和支持。此外，参与编写工作的还有刘庆芳、刘孟起、吴炳根、丁根秀、文桂秀、魏淑芹、张云铎、王铭东、黄文论、王真龙、刘倩、达其双、陈鑫、滕音。感谢电子工业出版社各位老师的辛勤工作，最后特别感谢永远关爱着我的家人。

本书的编写工作得到了国家自然科学基金（No. 61272049）、北京市自然科学基金（No. 4132075）的资助。

由于作者水平有限，书中难免存在纰漏，欢迎广大读者批评指正。在阅读过程中，如果发现问题，请发送 E-Mail 电子邮件告知，以便今后再版时加以修正。



吴怀宇

中国科学院

E-mail : huaiyuwu@gmail.com

主页网址 : <http://www.sigvc.org/people.htm#why>

于 北京中关村

目 录

第1章 3D打印与“全球第三次工业革命”	1
1.1 3D 打印：体验造物奇迹	1
1.2 全球第三次工业革命的导火索	10
1.2.1 从“第一次工业革命”到“第三次工业革命”	10
1.2.2 3D 打印的显著优势	11
1.2.3 3D 打印的应用现状	13
1.3 对 3D 打印的质疑	15
1.3.1 来自传统制造业大佬的质疑：不看好 3D 打印	15
1.3.2 关于“3D 打印技术的可实现性”释疑	16
1.3.3 关于“3D 打印技术的经济性”释疑	17
1.3.4 关于“3D 打印产业的成长性”释疑	18
1.4 3D 智能数字化与 3D 打印：用“虚拟”再造“现实”	20
1.4.1 3D 智能数字化设计技术的发展现状	20
1.4.2 智能数字化扫描技术的发展现状	21
1.4.3 智能云网：云端智能服务和云制造	22
1.4.4 3D 打印技术的发展现状	23
1.5 创客 DIY：新工业革命的启蒙运动	25
1.5.1 以小博大：创客挑战巨头公司	26
1.5.2 聚沙成塔：改变工业社会的组成结构	27
1.6 “中国制造”向“中国智造”转变的机遇	28
1.6.1 “中国制造”需转型升级	28
1.6.2 来自“德国制造”的启示	29
1.6.3 “中国智造”的发展机遇	30
第2章 3D打印机的原理与种类	37
2.1 3D 打印时间简史——源自 1860	37
2.2 3D 打印机的工作原理和家族	41
2.2.1 3D 打印机的工作原理与流程	41
2.2.2 FDM：熔融沉积成型（FFF：熔丝制造）	42
2.2.3 3DP：三维打印黏结成型（喷墨沉积）	44
2.2.4 SLS：选择性激光烧结	45
2.2.5 SLA：光固化立体成型（立体光刻）	47
2.2.6 PolyJet：多头喷射技术（Material Jetting：材料喷射）	49
2.2.7 DLP：数字光处理	50

2.2.8 LOM : 分层实体制造	51
2.3 塑料还是石膏? 3D 打印机的各种耗材	53
2.4 金属 3D 打印技术大盘点	59
2.4.1 SLS、SLM 和 DMLS 技术	60
2.4.2 LENS/LNSF/LPF/DMD/LC/DLF : 激光近净成型	63
2.4.3 EBM : 电子束熔炼	64
2.4.4 EBDM : 电子束直接制造	65
2.4.5 金属 3D 打印技术小结	66
2.5 两大阵营 : 工业级打印机与桌面级打印机	67
2.5.1 工业级打印机 : 两个巨头的主战场	67
2.5.2 桌面级打印机 : 创客们的多样世界	73
2.6 3D 打印与传统手办模型制作	78
2.7 3D 打印机购买指南	80
第3章 剖析3D打印机: 轮子是怎样发明的	83
3.1 RepRap : 开源 3D 打印机的鼻祖和奠基石	83
3.2 MakerBot 与 Ultimaker : 桌面双雄	84
3.3 Ultimaker 组装实战	86
3.3.1 Ultimaker 新到货开箱照	87
3.3.2 搭建框架	87
3.3.3 X/Y/Z 轴电机	90
3.3.4 X/Y 轴承	92
3.3.5 挤出头	94
3.3.6 Z 轴载物平台	98
3.3.7 送料机	101
3.3.8 Ultimaker 的大脑 : 电路板	103
3.3.9 大功告成 : 一台完整的打印机	106
3.3.10 Gcode 与前台软件 Cura 使用指南	108
3.3.11 Ultimaker 打印成果实例	113
3.4 MakerBot Replicator 2 与 MakerWare 打印实战	114
3.4.1 MakerWare 进行切片和打印	114
3.4.2 ReplicatorG 控制前台的设置 : 双喷头打印双色模型	120
3.4.3 MakerBot Replicator 2 打印成果实例	122
3.5 3D 打印疑问与故障排解小贴士	123
3.5.1 模型的水密性 (Watertight)	123
3.5.2 模型必须为流形 (Manifold)	123
3.5.3 切片 (Slice) 与横切面	125
3.5.4 层厚度 (Layer Thickness)	125

3.5.5 支撑材料 (Support Material).....	125
3.5.6 如何开始打印.....	125
3.5.7 如何调平打印平台 (粗调和精调).....	126
3.5.8 如何更换耗材 (上料、退料).....	126
3.5.9 我装不了塑料丝	126
3.5.10 我取不出塑料丝导管	127
3.5.11 为什么我的送料机挖坑，但就是不吐丝	127
3.5.12 喷头堵塞，如何处理	127
3.5.13 挤出的料无法粘牢打印平台	127
3.5.14 打印出的东西粘不牢平台	127
3.5.15 用辅助盘 (Helper Disks) 解决翘边问题	128
3.5.16 喷头位置偏移，挤出头坐标异常.....	129
3.5.17 为什么打印的圆是椭圆.....	129
3.5.18 电机不转，像得了帕金森症抖个不停.....	129
3.5.19 为需要连接的零件选择合适的容许公差	129
3.5.20 如何让模型表面更光滑.....	129
3.5.21 我的打印机需要日常维护吗	130
3.5.22 异常情况如何中断打印.....	130
3.5.23 如何将金属零件放入我的 3D 塑料模型中.....	130
3.5.24 用 CNC Simulator 进行打印模拟和打印预览	131
3.5.25 打印失败后是什么样子	131
第4章 3D智能数字化：3D打印的孪生兄弟	133
4.1 不以规矩，不成方圆——STL 数字标准文件解析	133
4.2 3D 智能数字化设计技术	136
4.2.1 “所想即所得”：3D 设计的新境界.....	136
4.2.2 商业设计软件：3D 设计的重型武器 (Maya、UG).....	140
4.2.3 杀鸡焉用牛刀：基于网页的设计软件 (Tinkercad、3DTin).....	146
4.3 3D 智能数字化扫描技术	148
4.3.1 光学三维扫描仪的原理和实例 (激光、结构白光).....	151
4.3.2 基于 Kinect 的 3D 扫描原理和设备 (红外光斑、ToF).....	156
4.3.3 房地产行业的新应用：室内 3D 扫描建模.....	162
4.4 面向“批量定制”和“柔性制造”的智能数字化.....	163
4.5 智能云网：云端智能服务和云制造	165
4.6 大数据和深度学习：3D 打印内容的挖掘与推荐.....	166
4.6.1 什么是大数据.....	166
4.6.2 大数据背景下的个性化推荐系统.....	168
4.6.3 深度学习：像人脑一样深层次地思考	171

第5章 3D智能数字化与3D照相馆：科学与艺术的结合	177
5.1 那些年，我们一起追过的 3D 照相馆.....	178
5.1.1 细数国内外的 3D 照相馆.....	178
5.1.2 3D 照相馆的设备与成本.....	180
5.1.3 3D 照相馆赢利模式的探讨.....	182
5.2 3D 照相馆的核心技术：3D 智能数字化.....	183
5.3 基于图像的 3D 人脸重建技术.....	186
5.3.1 基于单张照片的 3D 人脸重建及立体浮雕.....	187
5.3.2 基于多视角照片的 3D 人脸重建.....	189
5.3.3 人是种视觉动物：如何美化你的照片.....	194
5.4 Skanect：使用 Kinect 实现 3D 扫描.....	198
5.5 头发修补：3D 照相馆的头痛问题.....	201
5.5.1 使用 3D-Coat/ ZBrush 软件手工修补发型.....	201
5.5.2 基于视觉计算自动修补发型.....	206
5.5.3 Geomagic Studio：更通用的任意形状修补.....	209
5.6 3D 人脸表情形变与编辑.....	216
5.7 直接全彩打印，还是单色打印再上色.....	220
5.8 3D 打印数字化设计技巧.....	222
5.8.1 3DS Max 建模用于 3D 打印.....	222
5.8.2 Netfabb/Magics：修正你的 STL 打印文件.....	226
5.8.3 使用 AccuTrans 3D 转换 3D 文件格式.....	229
第6章 视觉计算：构建3D打印的杀手级应用	230
6.1 视觉计算：计算机视觉与计算机图形学的融合.....	230
6.2 3D 打印“批量定制”的智能实现.....	232
6.2.1 个性特征的描述与检测.....	233
6.2.2 个性特征的定位与匹配.....	237
6.2.3 个性化形状的编辑与合成.....	242
6.3 立体视觉重建：将照片转成 3D 数字模型.....	247
6.3.1 摄像机定标.....	247
6.3.2 基于立体视觉、SFM 和 Visual Hull 的三维重建.....	254
6.4 众里寻她千百度——海量 3D 模型的检索.....	257
6.4.1 线性分类与感知机模型.....	257
6.4.2 支持向量机 SVM 与逻辑回归 LR.....	260
6.4.3 基于内容的 3D 模型检索.....	263
6.5 形状拆解：大尺寸物件的自动分块打印.....	267
6.6 形状分析：优化桌面 3D 打印机打印精度的表现力.....	269
6.7 形状平衡：如何确保 3D 物件站立稳当.....	271

6.8 形状优化：生成坚固的内部轻质结构使得耗材最省	274
6.9 基于笔画的 3D 建模：让新手和孩子轻松设计形状	277
6.9.1 Doodle3D：3D 设计就像涂鸦一样简单	278
6.9.2 Teddy/FiberMesh：更精准的 3D 笔画建模	279
6.9.3 3-Sweep 技术：轻松让照片中的 2D 物体变 3D 模型	280
6.9.4 “神笔马良” 3Doodler：用笔直接画出 3D 线框实物	282
6.10 增强现实：在打印之前看到融入环境的真实效果	283
6.11 OpenCV 与 OpenGL：视觉计算入门的两大利器	284
6.11.1 OpenCV 与 AdaBoost 人脸检测	285
6.11.2 OpenGL 与 3D 图形绘制	290
第7章 创客：个人3D打印机的创造者	295
7.1 创客文化与开源 DIY	295
7.2 五花八门的创客杰作：从玩具到高速跑车	297
7.3 寓教于乐：3D 打印出你的个人数学博物馆	301
7.4 创客之开源硬件 Arduino（阿德伟诺）	305
7.4.1 Arduino 简介	305
7.4.2 初窥 Arduino	306
7.4.3 牛刀小试：叩开 Arduino 之门	308
7.5 创客之开源软件 Android（安卓）	310
7.5.1 Android 概述	310
7.5.2 开发平台搭建	311
7.5.3 Android 之旅起航：Hello, Android!	312
7.6 靠创意去赚钱：漫谈 Kickstarter、Quirky 与 Shapeways	316
7.6.1 Kickstarter 众筹：靠创意去筹资	316
7.6.2 Quirky 创意加工厂：把创意变成产品	317
7.6.3 Shapeways 在线打印：把个性化产品定制出来	319
7.7 创客中国：中国版乔布斯和比尔·盖茨的诞生地	320
7.7.1 国外创客为什么纷纷青睐中国	320
7.7.2 创客中国的背景优势	321
7.7.3 创客中国的市场细分定位	321
第8章 创客实战：四轴飞行器	323
8.1 你准备好了吗：自己制作四轴飞行器	323
8.2 器件与 3D 打印	324
8.2.1 四轴飞行器 DIY 所需的器件汇总	325
8.2.2 四轴飞行器的遥控器和接收机	326
8.2.3 四轴飞行器的飞行控制板	327
8.2.4 四轴飞行器电调的选用	328

8.2.5 四轴飞行器的无刷电机和螺旋桨.....	329
8.2.6 四轴飞行器的电池和充电器	330
8.2.7 四轴飞行器的连接线选用.....	331
8.2.8 四轴飞行器机架的 3D 打印	331
8.3 三轴陀螺仪和加速度计的入门与调试	332
8.4 自制基于 Arduino 的飞控板	335
8.4.1 四轴飞行器的基本电控结构	335
8.4.2 飞行控制板的制作	337
8.5 遥控开始：Android 手机的 Wi-Fi 通信	339
8.6 四轴飞行器的智能视觉跟踪.....	341
8.6.1 基于粒子滤波的目标跟踪算法	342
8.6.2 基于 Mean Shift（均值漂移）的目标跟踪算法.....	345
第9章 3D打印之不久的将来.....	348
9.1 3D 打印的未来：由创客们决定	348
9.1.1 几乎为零的设计和制造门槛	349
9.1.2 创客成就 3D 打印	349
9.2 手机应用 FabApp、App Store 与智能云网	351
9.3 不再仅仅是看着粗糙的 FDM	352
9.4 生物医疗打印：越来越近的科幻	353
9.5 美食打印机：“吃货”的钱最好赚	356
9.6 绿色经济：变沙漠为光影城市	359
9.7 打印房屋：安得广厦千万间.....	361
9.8 混合材料制造：3D 打印电路.....	363
9.9 枪支打印“让子弹飞”、版权与社会伦理	365
9.9.1 3D 打印引发社会公共安全的忧虑.....	365
9.9.2 版权保护的难题	367
9.9.3 社会伦理的思考及技术层面解决.....	369
9.10 3D 打印 3D 打印机自己：遗传与升级.....	370
9.11 3D 打印的经济模式：利基与长尾效应.....	371
9.12 “中国智造”推动“全球第三次工业革命”	374
9.12.1 新工业革命之“永不枯竭的绿色能源”	374
9.12.2 新工业革命之“3D 打印新材料”	375
9.12.3 新工业革命之“先进制造及 3D 打印”	376
9.12.4 新工业革命之“3D 智能数字化创造”	377
第10章 道：数字智能的最优化及相关数学方法	379
10.1 最优化理论的基本常识	380
10.1.1 从凸集和凸函数开始说起	380

10.1.2 无约束优化与约束优化.....	383
10.1.3 线性规划与非线性规划及其对偶（Dual）形式.....	383
10.1.4 澄清混淆：二次规划、二次收敛、二阶收敛.....	385
10.2 最优化根基之单变量“一维搜索”.....	385
10.2.1 初始搜索区域的加步探索法（进退法）.....	386
10.2.2 黄金分割搜索法（Golden Section Search）.....	386
10.2.3 斐波那契（Fibonacci）搜索法.....	387
10.2.4 牛顿法、抛物线法.....	388
10.2.5 不精确线搜索的 Armijo-Goldstein 准则及 Wolfe-Powell 准则.....	389
10.3 多变量的无约束优化.....	391
10.3.1 最速下降法（Steepest Descent，梯度下降法 Gradient Descent）.....	391
10.3.2 牛顿法（Newton）.....	392
10.3.3 拟牛顿法（Quasi-Newton）：DFP 和 BFGS 方法.....	393
10.3.4 共轭方向法（Conjugate Direction）.....	394
10.3.5 共轭梯度法（Conjugate Gradient）.....	395
10.3.6 Powell 直接法.....	396
10.4 最优化根基之“信赖域”.....	396
10.4.1 Levenberg-Marquardt（L-M）方法.....	398
10.4.2 详解 L-M 方法的求解过程与步骤.....	398
10.5 最小二乘问题的求解.....	399
10.5.1 线性最小二乘问题的求解（正规化方法、QR 分解、SVD 分解）.....	399
10.5.2 非线性最小二乘问题（Gauss-Newton 方法）.....	400
10.6 约束优化问题的求解.....	401
10.6.1 等式约束的拉格朗日乘子法（Lagrange Multiplier）.....	401
10.6.2 不等式约束的 KKT（KT）条件.....	401
10.6.3 惩罚函数法（外点法、内点法）.....	402
10.7 最短路径与动态规划（Dynamic Programming）.....	403
10.8 “偶然中的必然”——概率与贝叶斯（Bayes）.....	404
10.8.1 先验概率、似然函数、后验概率、贝叶斯公式.....	404
10.8.2 朴素（Naïve）贝叶斯分类.....	405
10.8.3 最大似然估计、最大后验概率估计、贝叶斯估计.....	407
10.8.4 贝叶斯学派与频率学派之争论.....	410
参考文献.....	412
后 记.....	416
作者简介.....	418

第1章

3D打印与“全球第三次工业革命”

《老子》(六十四章)有云：“合抱之木，生于毫末；九层之台，起于累土。”以两千年前思想大家老子的这句话作为全书的开篇，应是再合适不过了。首先，它几乎准确地“预言”了3D打印的过程：从“毫”米级（如标准的1.75mm、3.0mm）的细丝或更细的粉“末”开始，通过逐层“累”积的方式，制造出3D物体。其次，以3D打印机这么小小的一台机器，谁又能想到它竟被英国著名经济学杂志《经济学人》声称将引发史诗般宏大的全球第三次工业革命呢？通过老子的这句深含哲理的话，我们可以领悟到：不要小看了事物的毫末之始，新工业革命的“九层之台”正由它而立起！

在本章中，我们从3D打印“神话般”的造物体验说起，也对目前社会上关于3D打印的一些质疑进行了探讨。实际上，如果只把眼光局限于3D打印目前的产能，让它与大规模流水线上的“大块头”工业机器们比速度、比谁的力气大、比谁干的活儿多，就着实“大材小用”了。3D打印与智能数字化是一对孪生兄弟，如果没有了智能数字化这个兄弟，3D打印也确实只能去拼力气、比活儿多了。而当有了智能数字化，结合创客们的创新精神，“中国制造”必将抓住“中国智造”的历史机遇，给中华文明谱写崭新的篇章。

1.1 3D打印：体验造物奇迹

近年来，我们经常能听到“3D”这个名词，且往往跟高科技联系在一起，如3D显示、3D电影、3D扫描、3D打印等。按理说，人类每天就生活在三维空间中，3D对我们来说本应是一个再寻常不过的概念。3D之所以被认为是“高科技”，很大程度上归因于我们通过高科技的数字化手段，使得客观世界中的3D实体能够在虚拟世界中得以高精度重建（3D扫描）、智能化编辑（3D设计）、真实感高清展示（3D显示），乃至重新返回至客观世界（3D打印）。就学科专业而言，3D技术横跨计算机视觉、计算机图形学、模式识别与智能系统、复杂系统与自动控制、数据挖掘与机器学习、工程材料学、光机电一体化等，是名副其实的“技术密集型”高科技。

3D打印（三维打印）是**增材制造技术（AM, Additive Manufacturing）**的俗称。**3D打印，名为“打印”，实为“制造”，结合智能数字化，更可实现“创造”！**实际上，在大量的英文文献中，

3D 打印（3D Printing）常被称作 3D Fabrication（3D 制造），这更准确地描述了 3D 打印的本质。

与传统的“切削去除材料”的加工技术（如 3D 雕刻）完全不同，3D 打印以经过智能化处理后的 3D 数字模型文件为基础，运用粉末状金属或塑料等可热熔黏合材料，通过分层加工、叠加成型的方式“逐层增加材料”来生成 3D 实体。由于可采用各种各样的材料（液体、粉末、塑料丝、金属、沙子、纸张，甚至巧克力、人体干细胞等），而且可以自由成型（任意复杂的中空镶嵌形状），所以 3D 打印机是名副其实的“**万能制造机**”。

图 1-1 是一幅关于 3D 打印的漫画，一个人把自己给打印了出来，相信一定会让你印象深刻。接下来，就让我们一起体验 3D 打印带来的造物奇迹吧！



图 1-1 一幅关于 3D 打印的漫画，一个人打印出了自己（图片来源：《经济学人》）

如图 1-2 所示的这些雕塑是由 Bathsheba Grossman 用复杂的激光烧结而成的。对于要求具有复杂的内部中空、凹陷、互锁或者有大量规则细节图案的形状加工，3D 打印机是首选的制造设备。智能数字化设计（图 1-2 左 1）可对零件进行优化，减轻重量，同时保持原有强度和其他关键性能。由于是一次成型，使产品成为一个整体，这样也**减少（或免除）零部件的装配**。



图 1-2 3D 打印出的任意复杂设计形状，传统制造工艺无法加工（图片来源：Bathsheba Sculpture）

如图 1-3 所示，3D 打印的衣服和鞋子亮相巴黎时装周。仔细查看衣服上无数精细花纹，我们可以知道：精美的手工刺绣和针线活已被智能数字化和 3D 打印取而代之。

在材料上，服装师们也已经有了更多、更好的选择，比如 Materialise 公司最新研发的 TPU92A-1 合成材料，可用于设计很酷的时装：这种材料拥有非常好的弹性和韧性，能够在受到高强外力的情况下快速恢复原来形状。

3D 打印不仅可以用来制造独具个性的高跟鞋（图 1-3 右上角），还可以帮助设计师实现更多灵感。图 1-3 右下角的这双鞋的设计灵感来自大自然中的树根，3D 打印技术能够建立错综复杂、相互交织的须状鞋跟，模仿树根在脚下扭曲和盘旋。这是其他任何一项制造技术都做不到的。



图 1-3 3D 打印的衣服和鞋子闪耀亮相巴黎时装周（图片来源：Reuters）

耐克（Nike）“Vapor Laser Talon”跑鞋，如图 1-4 所示，抛弃了传统制鞋工艺，使用 3D 打印制造，只有 158.7g（相当于 3 枚鸡蛋的重量），可以帮助运动员在草地赛场上“飞”起来。这款鞋子仅在 7 个月之内，就已经试做、试用了成百上千双，直到最终满意定型。传统工艺在如此短的时间内根本无法做到，只能通过 3D 打印来实现。



图 1-4 耐克的定制跑鞋，仅 3 枚鸡蛋重，让运动员找到“飞一样”的感觉（图片来源：Nike）

3D 打印的 OpenReflex 胶片单反相机（如图 1-5 所示），成本相当便宜（不到 200 元人民币），只需要 15 小时的 3D 打印及 1 小时的组装就能使用了。这台相机虽然只能使用 1/60s 快门，却

可以兼容任何镜头，可谓相当强悍，其主要部件都是用 ABS 材料打印的。

俄罗斯大皇冠是为 1762 年凯瑟琳的加冕礼打造的。它作为官方权力的象征，每任君王都需要戴着它。通过智能数字化扫描和设计，3D 打印机已精确复制出原版皇冠，如图 1-6 所示。复制品采用 WIC-100 树脂以及 14K 白金打造。由于使用了 3D 打印机，将制造时间由预计的一年缩短到两个月，而且所达到的精准度超越了手艺人采用的常规方式。



图 1-5 3D 打印的胶片单反相机，成本不到 200 元（图片来源：Thingiverse）



图 1-6 沙俄大皇冠完美复制品采用 3D 打印珠宝钻石（图片来源：Envisiontec）

维也纳科技大学推出的“纳米级 3D 打印机”可打印出纳米级的物件，让很多人惊叹的微雕作品现在就可以通过这台打印机轻松搞定。如图 1-7 所示，我们看到的是 3D 打印出的维也纳史蒂芬大教堂和 F1 赛车模型，直径不超过人类的头发丝，比一粒沙子还小。纳米级 3D 打印机使用液态树脂，通过激光使树脂硬化成型，此技术被称为**双光子光刻（2PP，双光子聚合、二光子平板印刷 TPL）**，而且打印速度非常快，可达 5m/s，创造出了一个新的世界纪录。

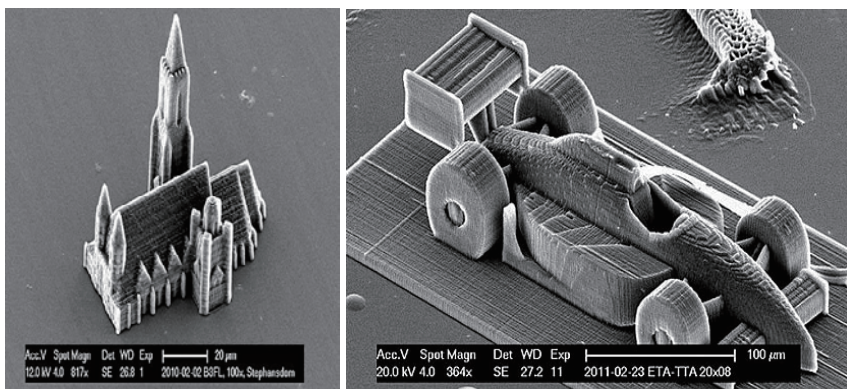


图 1-7 纳米级 3D 打印问世，实现微雕制作（图片来源：维也纳科技大学）

美国国家航空航天局（NASA）宇航员前往火星需要携带大量食物，3D 打印机可以按照“数字菜谱”混合各种粉末，制造色香味俱全且营养可口的太空食品。食物原料从藻类、甜菜叶及昆虫等可持续资源中提取，经过脱水处理后，保质期可延至 30 年左右。这个计划很快就会得以实现，不信？请看看目前 3D 打印机所打印出来的各种饼干、巧克力和糖品，如图 1-8 所示，诱人否？



图 1-8 3D 打印机所打印出来的各种饼干、巧克力和糖品 (图片来源 : Evil Mad Scientist Lab)

荷兰建筑师 Janjaap Ruijsseenaars 所主导的 3D 打印项目如图 1-9 所示,这将是全球首个 3D 打印的大型建筑物。他的计划是打印 $6\text{m} \times 9\text{m}$ 的模块框架,这些模块由沙子和无机黏结剂组成,然后使用钢纤维混凝土填充,最终的产品将是一幢二层楼。这是一个“莫比乌斯带”(Möbius Band) 式的建筑(它没有正反面之分,即曲面是不可定向的 Non-Orientable,在上面一直走,可一次走完所有面)。由于采用了创新的 3D 技术进行打造,因此它将成为一座兼具延伸性和适用性的建筑。与此同时,麻省理工学院(MIT)的研究人员正在开展另一项研究,如何利用 3D 打印在一天之内打印出房屋主要结构,而利用传统的建筑团队则需要一个月的时间才能完成。



图 1-9 荷兰建筑师设计的“莫比乌斯环”建筑,将采用 3D 打印进行建造
(图片来源 : Janjaap Ruijsseenaars)

如图 1-10 所示,美国研究人员花了 4 个月设计时间,3D 打印出无人飞机,巡航时速可达到 45 英里,成本仅 2 000 美元。该飞机翼展约 2m,所有零部件都是通过 3D 打印机制造出来的。经打磨喷漆处理之后,外形也非常时尚。而在 5 年前,光制造塑料涡扇发动机的成本就约为 25 万美元,还需要花上两年时间。



图 1-10 3D 打印出的无人飞机，巡航时速 45 英里，成本仅 2 000 美元（图片来源：弗吉尼亚大学）

美国重型机车厂 OCC,利用大型的商用 3D 打印机成功打造出一款咆哮的“中国龙”型摩托车，如图 1-11 所示。该车由一位中国客户定制，龙头部分由 3DS Max 软件设计后，输出到 3D 打印机一次打印成型，然后直接组装和整合到机车上就可以了，大大简化了龙头的设计和和生产周期，同时也极大地降低了制造成本。劳斯莱斯，纯手工打造？可以肯定的是，手工打造以后将不再显得那么金贵。



图 1-11 3D 打印出的一款咆哮的“中国龙”型摩托车（图片来源：OCC）

随着外科修复科技的日益发展与手术理念的日渐人性化，假肢不仅很舒适而且也可以很时尚！如图 1-12 所示，由于每个人的身材和喜好不同，所以需要进行完全个性化的定制。首先使用 3D 扫描仪取得用户腿部详细数据，然后根据用户自身数据、年龄、性别、特殊要求等，设计出适合的假肢款式，用户觉得满意后，再通过 3D 打印机进行制造。



图 1-12 3D 打印出的时尚假肢，根据用户身体和喜好完全个性化定制（图片来源：3D Systems）

如图 1-13 所示的这只美洲雕 2005 年曾被误伤，从此生活不能自理。如今，通过 3D 扫描鹰喙的残缺部分并使用 SolidWorks 软件对数字化模型进行编辑处理，然后再用尼龙聚合物材料将假体 3D 打印出来。于是，这只海雕得以顺利回归了大自然。从图片上可以看到，鹰嘴具有精细的内部中空结构，而 3D 打印的再造能力丝毫不亚于最好的整形医生。



图 1-13 3D 打印帮助美洲雕再造完美的鹰嘴(图片来源 : Nate Calvin)

之前已有报道 3D 打印可制造出真正有触觉的生物人耳（不只是装饰性的模型）。现在，普林斯顿大学的研究人员利用 3D 打印的细胞和纳米粒子，结合小型线圈天线软骨组织创造了一只仿生耳，如图 1-14 所示，竟能够听到超越人类所能听到的无线频率！以后，生物打印机将使用病人自身的干细胞，那么器官移植后的排异反应将会减小。随着整个世界逐渐变得越来越电子化和数字化，我们将制造出一些新型器官，以便和手机、便携式电脑进行直接交流。



图 1-14 3D 打印的硅树脂仿生耳，人类未来可获“第六感”(图片来源 : 普林斯顿大学)

美国国家航空航天局(NASA)使用直接金属激光烧结的 3D 打印技术制作了火箭喷射器部件，如图 1-15 左边所示。据悉可承受 1 400 磅每平方英尺的压力并可耐 6 000 华氏度的高温。而且该喷射器在削减成本方面迈出了一大步，因为它仅由 2 个零件组成，而此前传统技术制造的同类型喷射器由 115 个零件组成。而更让人唏嘘的是，美国加州大学的学生也用 3D 打印制造了一种微型火箭进行了成功的点火测试，如图 1-15 右边所示，他们在短短 8 个月内就完成了火箭金属发动机的设计，且火箭的全部制造费用只有 6 800 美元。

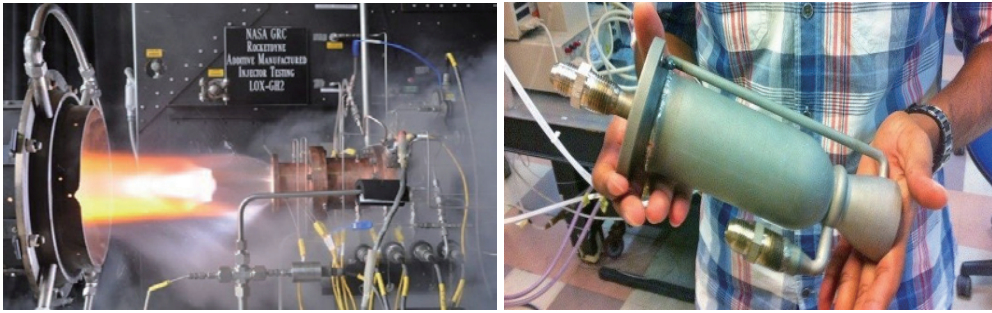


图 1-15 左：NASA 利用 3D 打印技术制作的火箭喷射器部件；右：学生也用 3D 打印制作微型火箭
(图片来源：NASA、UCSD)

3D 打印技术集概念设计、技术验证与生产制造于一体，这必将极大缩小武器装备从“概念”到“定型”的时间差，从而加快武器装备的更新周期，如图 1-16 所示。据媒体报道，歼-10 战斗机研发用了近 10 年时间，而运用 3D 打印技术后，我国仅用 3 年就推出了第一款舰载机歼-15。目前，3D 打印技术已被全面应用于第一款本土隐形战斗机歼-20 和第五代战斗机歼-31 的研发中。不容置疑，3D 打印技术正在制造空军发展的“中国速度”。



图 1-16 3D 打印技术用于战斗机研发 (图片来源：vx.com)

再说个喜庆点的案例。你完全可以自己在计算机上设计戒指的形状，打印出专属的订婚戒指，如图 1-17 所示。拿着这样的戒指求婚，相信没有一个女孩会说不。



图 1-17 3D 打印的戒指和吊坠 (图片来源：Shapeways)

目前在北京、西安、武汉等地都相继开设了3D照相打印馆。本书第5章“3D智能数字化与3D照相馆：科学与艺术的结合”将从零起步、手把手地教你开设3D照相馆，包括3D人像扫描和3D打印塑像，如图1-18所示。



图 1-18 从左至右：笔者本人的照片、3D 数字化模型；3D 打印后的头像

打印了3D头像，再打印个3D头盔吧！国外一名叫 Ryan Brooks 的创客，用塑料打印出了钢铁侠头盔，如图1-19所示。整个开发和微调过程大约花了100小时。头盔采用 Adafruit 加速度传感器和 Arduino Pro 迷你伺服机构，可自动开启或关闭头盔的面罩。



图 1-19 国外创客打印的钢铁侠头盔（图片来源：Ryan Brooks）

最后再介绍“宅男”们的最爱——人体模型制作。如图1-20上方所示，日本的DIY爱好者借助3D数字化技术和3D打印技术制作了动漫形象的立体人偶。而如图1-20下方所示，日本东京一家公司利用3D打印技术，对新娘进行克隆，制作出仿真模型，以留住她们人生当中最美好的时刻。相信在未来，这个方向将大有可为。



图 1-20 3D 打印技术进行仿真人体模型制作(图片来源 : Danny Choo)

通过以上的一番体验，你是不是开始感觉到**3D 打印已经渗透进我们生活中的“衣食住行”**等各个方面了呢？实际上，3D 打印可广泛用于**工业制造、珠宝首饰、玩具设计、机器人、生物医学、建筑与城市规划、食品制作、航空航天、考古科研**等领域。以音乐乐器为例，目前大多数乐器，如吉他、小提琴、喇叭、钢琴等，已经有几百年都没再改进过。而使用 3D 打印，你不仅可以自由设计出奇形怪状的创意乐器（比如将一把马头琴变成印有你自己 3D 头像的琴），而且更重要地，你可以**发挥 3D 打印轻松制造任意复杂空腔（共鸣腔）的优势**，设计制造出一把发声效果最佳的、属于我们 21 世纪的新型乐器，演奏属于我们这个全新智造时代的新乐章！

1.2 全球第三次工业革命的导火索

3D 打印跨越了虚拟的比特世界和实体的原子世界之间的鸿沟，其革命性的意义超越了之前个人电脑和互联网的出现。2012 年，《经济学人》、《福布斯》、《纽约时报》等杂志都称 3D 打印将引发“第三次工业革命”，期望以此让制造业重新回流到欧美等西方发达国家。据预测，3D 打印行业的产值将在 2016 年达到 31 亿美元。2012 年 8 月，美国总统奥巴马拨款 3 000 万美元，在俄亥俄州建立了国家级 3D 打印添加剂工业研究中心，并计划第一步投入 5 亿美元用于 3D 打印，以确保美国制造业不再继续转移到中国和印度。但笔者在本章第 1.6 节中认为恰恰相反，3D 打印相关技术将给新兴国家带来更多机遇，将使制造业尤其是制造业的上游产业链，进一步掌握在中国等新兴国家手中。

1.2.1 从“第一次工业革命”到“第三次工业革命”

第一次工业革命于 18 世纪从英国发起，它开创了以机器代替手工劳动的时代，典型特征为“**机械化**”。18 世纪 60 年代，珍妮纺纱机（1765 年）的发明和应用是第一次工业革命开始的标志。瓦特（1781 年）改良的蒸汽机，将人类带入了“蒸汽时代”。因此，蒸汽机是第一次工业革命的

主要标志。同时，英国也因为引领了第一次工业革命而建立了“日不落帝国”。

第二次工业革命起始于 19 世纪 70 年代，它以电力的大规模应用为代表，典型特征为“**自动化**”。1866 年德国人西门子制成发电机，电力开始用于带动机器，成为补充和取代蒸汽动力的新能源。电灯的发明以及电力工业和电器制造业迅速发展，人类跨入了“电气时代”。20 世纪初，美国福特汽车公司大规模生产流水线的诞生成为了第二次工业革命的重要标志。同时，德国和美国也因为引领了第二次工业革命而相继崛起乃至倨傲群雄。

第三次工业革命是以智能数字化制造及新型材料应用为代表的的一个崭新的时代，具体特点可描述为：智能数字化、分布式网络化、个性定制化、绿色可持续化，典型特征为“**智能数字化**”。20 世纪 80 年代美国 IBM 公司推出世界上第一台个人电脑，20 世纪 90 年代互联网开始大规模应用于商业领域，人类从此进入了“信息时代”。21 世纪初 3D 打印的普及将引爆第三次工业革命。3D 打印、智能数字化、新材料，以及机器人技术的发展，将极大地改变制造业原有的投入模式，使得依靠较少的自然资源和人力资源投入，就能取得良好经济效益成为可能。可以预见，未来的生产制造将更有**柔性**（即**灵活性**），将远离产品千篇一律的大规模制造模式，向更具个性化的生产模式发展。

同时，根据前两次工业革命的历史经验，比如英国、德国、美国的崛起历史，我们是不是也可以认为，谁引领了第三次工业革命，谁就将实现全面的复兴和崛起？

1.2.2 3D 打印的显著优势

3D 打印之所以具有革命性的意义，主要是集**两大突出优势**于一身。1) 个人只需在计算机中进行智能化设计，然后将复杂作业流程转化为数字化文件，发送到 3D 打印机即可实现制造。整个过程中，用户根本无须掌握各种复杂的制造工艺和加工技能，这样大幅降低了制造业的技术门槛。2) 由于 3D 打印的逐层加工、累积成型的特点，制造几乎不受结构复杂度的限制，结合智能数字化设计，可轻松实现产品的个性化定制。

过去各种实体产品的生产都必须依赖大型工厂的昂贵专业设备。随着在 2008 年至 2009 年间 3D 打印市场发生了重大转折，RepRap 开源打印机以及各种衍生低价产品的出现（低于 1 000 美元），使得个人也完全可以拥有一台 3D 打印机。3D 打印可以让人们只需动动鼠标就生产出各种东西，这大大缩减了一个想法到一件真实产品原型的距离。3D 打印将虚拟世界的数字内容和物理世界的实体物品连接了起来。

对于个人创业而言，以前有两个门槛：创意和技能。但是有了 3D 打印技术之后，技能这个门槛已经越来越低。创业者只需要在计算机里做出创意的原型，然后不需要工厂的帮助（原本需要给工厂支付一大笔钱，比如开模的费用）就能用 3D 打印机做出原型。3D 打印给了普通人以制造的能力，释放个体使用者的创新冲动，改变了过去发明创造只是少数人的特权。3D 打印提前介入创意设计环节，让设计者能够白天设计，晚上用 3D 打印机打印出来，次日可以进行讨论修改，如此不断调整，最终的产品能够更好地体现设计者的构想和满足用户的需求。

如图 1-21 左边所示，上大学时有过金工实习经历的读者会很清楚，这堆复杂的齿轮加工起来非常困难，而且它们之间的啮合还是活动的，而不是死的，这就要求特别高的加工精度。传统

技术需要你去了解很多知识和技能，如使用数控机床做加工，你需要学习好多年才能成为一名熟练的操作工（实际上，即使你学徒 8 年也无法加工这个整体一次成型的齿轮组，因为只有 3D 打印才能做到）。作为对比，3D 打印机的操作很快就可以上手，以我们的经验，快的人 10 分钟，慢的人 1 小时就可以学会。还有人会说 3D 打印只能制造一些小玩意，那请看看图 1-21 右边的这个大扳手！而且这个扳手还是个可调节开口宽度的活扳手哦。齿轮和扳手都是一次成型的，因此都无须任何组装。

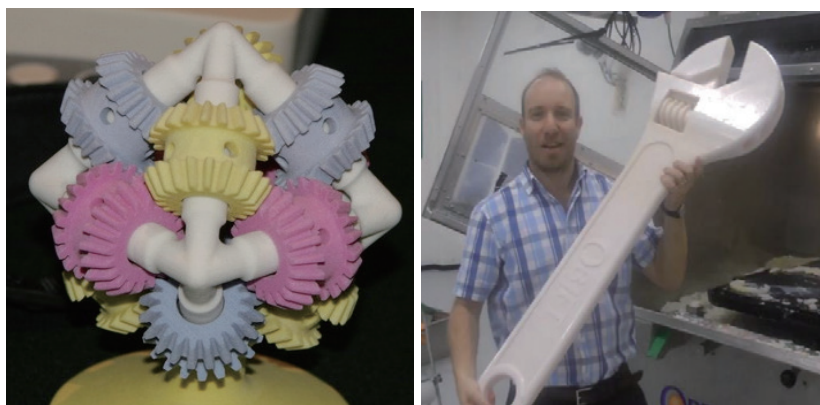


图 1-21 3D 打印的滚动齿轮和可调节扳手，无须组装（图片来源：Zach Walton）

有的读者会说，上面的齿轮和扳手都是塑料做的，结实吗？实际上，塑料在许多制造业内取代了金属，它的优势是重量轻、耐腐蚀、可塑性好等。如今越来越多的新型塑料研制成功，不仅具有非常高的强度，而且有的还具有很强的弹性和记忆属性，成本也都不高。当然，除了新型塑料，还有越来越多的新材料被研发出来，将掀起一股新的 3D 应用风潮。来自密西根理工大学的研究员 Joshua Pearce，就对 3D 打印未来的普及程度进行了大胆预测，随着 3D 打印机和材料的廉价化、实用化和普及化，以 3D 打印为代表的个人智造将会像个人电脑一样，很快成为主流。

此外，除了刚才介绍的两个突出特点，3D 打印技术还具有如下几大优势。

性能卓越

基于 3D 打印技术，设计人员可以对产品的内部结构进行精细控制以获得最佳效果。例如，用晶格或蜂窝状内部结构取代一个整块，可以在减轻产品重量的同时又不牺牲强度。

此外，研究人员正在研究一系列技术来控制打印件的性能，甚至能对金属的微观晶体结构精细控制，这将在本质上改变材料的底层原子和分子排列。例如，传统的金属铸锻技术（即**受压成型**）需要金属从外至内冷却，而金属 3D 打印采用快速凝固，从而导致更均匀的微观结构。因此，工程师可以控制成品的强度、硬度、弹性、灵活性和耐压力。

就某些材料而言，3D 打印不只是较好的选择，更是理想的生产方式。钛就是一个例子——重量轻、强度（密度）比钢强、比不锈钢更耐腐蚀。事实上，在许多应用场合，钛都是近乎完美的金属选择。然而，除了成本较高外，钛的主要缺点是：在切割过程中容易硬化，这导致刀具磨损严重；在焊接过程中，又容易受到污染，这导致焊接点容易脱落。而 3D 打印技术却可以很好地对钛进行驾驭——因为此时的钛已成为了一堆很细的粉末，只需不断地添加烧结即可，不

存在任何加工问题，既不需要切割也不需要焊接。

成本优势

以前你要找个工厂，让它为你生产一把你自己设计的锤子，首先你至少要为此支付 5~10 万元的开模费用。如果产量低于一百，则单件的价格高达几百元。因此，如果仅做一把或几把锤子，成本将是无比高昂的。但是，对 3D 打印机而言，无论是生产一件产品还是一千件产品，设备成本都是一样的。

此外，3D 打印和传统制造业的另一个区别在于产品的形成过程。传统制造过程通常使用消减的做法，包括研磨、锻造、弯曲、成型、切割、铣削、焊接、黏接、装配等。整个过程中会浪费很多原材料，同时在金属加热和再加热的过程中产生大量的能源消耗。

与此相反，3D 打印技术在小批量打印方面已经表现出了显著的价值。首先无须采购各式各样的机床，如车床、铣床、磨床等，这就省去了一大笔设备采购、维护费用。同时，因为是有控制地一层层添加材料，加工废料也大大减少，可以留下 90% 的原材料。以国外某个制造厂为例，通过使用 3D 打印（采用熔融沉积成型方法，即 FDM）定制注塑模型的某个特别部件，制造成本由 10 000 美元降至 600 美元，生产时间从 4 周减少到 24 小时，且重量减轻了 70%~90%。

更重要的是，3D 打印技术在样件设计制造上优势明显，省去模具制造的过程，在提升研发速度的同时，降低了研发失败的成本。当然，3D 打印技术也可作为大规模生产的辅助工具，比如模具和其他工具的制造，用传统加工方法一般需要花费一个月的时间，而使用 3D 打印在 48 小时内就可以完成。

在产品直接制造方面，比如使用 3D 打印实现多层电路一次成型的整合制造，可具有明显的速度优势。

更重要的是，3D 打印具有“**即需即印**”的优势，当顾客下单后，定位一个距离顾客物理位置最近的云制造节点开始制造，然后迅速送货上门，这样省去了产品库存、物流的成本。

1.2.3 3D 打印的应用现状

前面介绍了 3D 打印的诸多优点，正是这些优点才点燃了全球第三次工业革命的导火索。然而，3D 打印的应用刚刚开始全面普及，肯定也存有很多不完善的地方，比如制造精度相对较低、制造简单结构部件的速度较慢等。如图 1-22 所示给出了 3D 打印技术的优势和劣势。

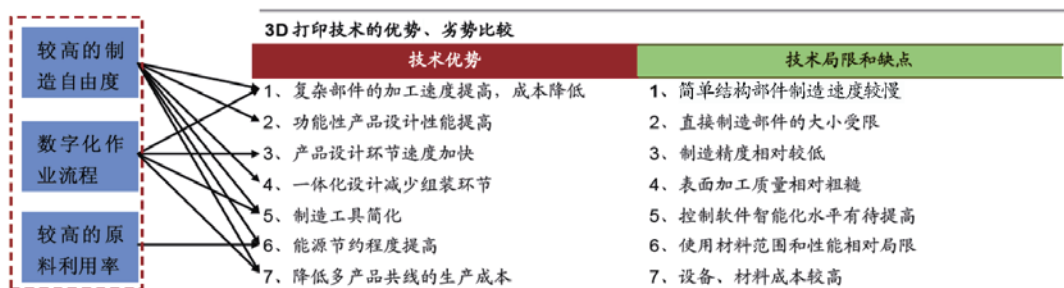


图 1-22 对 3D 打印技术进行优势、劣势比较分析（来源：华泰证券的行业调研）

从图 1-22 可以发现，当前的 3D 打印技术其实更适合于个性化定制需求较多、产品更新换代较快的市场环境，可使得从设计到推向市场的时间（包括样件制造、实验测试、模具制造）大幅缩短。3D 打印将会越来越广泛地应用于产品开发设计阶段的原型（样件）制作、辅助工具制造（如模具）、直接生产高度定制或技术复杂的小批量产品。

3D 打印需要依托多个学科领域的尖端技术，至少包括信息技术、精密机械和材料科学三大技术。通过与数控加工、铸造、金属冷喷涂、硅胶模等制造手段结合，在航空航天、汽车摩托车、家电、生物医学等领域得到了一定的应用，在工程和教学研究等应用领域也占有独特地位。具体应用领域至少包括以下几个。

- 生物医疗：用于制作人造骨骼、牙齿、助听器、假肢等。
- 航空航天、国防军工：用于直接制造复杂形状、微细尺寸、特殊性能的零部件。
- 消费品：珠宝、服饰、鞋类、玩具、创意 DIY 作品的设计和制造。
- 文化创意和数码娱乐：通过设计形状和结构复杂、材料特殊的作品来进行艺术表达。3D 打印的小提琴已接近手工艺的水平。
- 工业制造：用于产品概念设计、原型制作、产品评审、功能验证；制作模具原型、直接打印模具，甚至直接打印产品。3D 打印的小型无人机、小型汽车等概念产品已问世。3D 打印的家用器具模型，也被用于企业的宣传、营销活动中。
- 建筑工程：建筑模型风动实验和效果展示、建筑工程的施工模拟。
- 教育：打印模型来验证科学假设，用于不同学科的实验和教学。在北美的一些中学、普通高校和军事院校，3D 打印已经被用于教学和科研。

如图 1-23 所示给出了 3D 打印具体应用于各个行业时的优势发挥和局限程度。

3D 打印技术的优势和缺陷对下游行业的影响				
	医疗行业	航空航天	消费电子	汽车设计、制造
技术优势	1 复杂部件的加工速度提高，成本降低	✓✓	✓	✓
	2 功能性产品设计性能提高	✓✓	✓✓✓	✓✓
	3 产品设计环节速度加快	✓	✓✓	✓✓✓
	4 一体化设计减少组装环节	✓	✓✓	✓✓
	5 制造工具简化	✓✓	✓✓	✓
	6 能源节约程度提高		✓✓	✓
	7 降低多产品共线的生产成本	✓✓✓	✓	✓
技术局限和缺点	1 简单结构部件制造速度较慢		×	×
	2 直接制造部件的大小受限	×		×
	3 制造精度相对较低	×	×	×
	4 表面加工质量相对粗糙	×	×	×
	5 控制软件智能化水平有待提高	×	×	×
	6 使用材料范围和性能相对局限	×	×	×
	7 设备、材料成本较高	×	×	×

图 1-23 3D 打印具体应用于各个行业时的优势发挥和局限程度（来源：华泰证券的行业调研）

新产品的原型制造是目前 3D 打印最主要的商业应用，约占 70% 的 3D 打印市场。原型使设计师（和他们的客户）可以在设计阶段早期触摸和测试设计理念或功能实现，从而避免了后续变更造成的昂贵代价，为新产品上市节省了大量的时间和金钱。以赤石（Akaishi）——日本的一家保健鞋和按摩设备制造商为例。该公司发现，通过 3D 打印原型，新产品从订货到交货的时

间缩短了 90%，并且使设计师在产品上市前就对功能有 100% 的信心。原型还有利于实验和创新，例如，使用 3D 打印技术，贝尔直升机公司可以在数天内完成新设计的测试，而使用传统方式需要花上数周。

在某些行业中，3D 打印已经从原型制造发展为直接零件生产，也称为直接数字化制造。EOIR 技术公司是一家领先的防御系统设计和开发公司，使用 3D 打印机制造坚固耐用的坦克外置设备。自从引入 3D 打印技术后，该公司的制造成本从原来的单件 10 万美元以上，下降到如今的 40 000 美元以下。再例如，在航空航天领域，空中客车通过 3D 打印来制造金属机翼支架，如图 1-24 所示，由于 3D 打印可以毫不费力地制造内部任意复杂中空的结构，使得部件重量较轻，飞机的重量也随之减轻，从而节省了燃料。



图 1-24 空中客车 3D 打印的金属机翼支架，比传统机翼支架更轻盈（图片来源：CSC）

3D 打印是一种数字化技术，而不仅仅是一种制造技术。它所具有的开放性和大众性，为创新搭建了舞台。它使制造业的门槛降低，点燃了从公司到大众的创造力，为伟大的全球第三次工业革命的来临铺平了道路。

1.3 对3D打印的质疑

尽管当前 3D 打印技术在全球范围内迅猛发展，甚至被誉为是第三次工业革命，不过郭台铭却公开表示，“如果 3D 打印真的能颠覆产业，那我的‘郭’字倒过来写”。

1.3.1 来自传统制造业大佬的质疑：不看好 3D 打印

据台湾《联合晚报》报道，制造大佬、鸿海董事长、富士康公司总裁郭台铭公开表示，3D 打印绝不等于第三次工业革命，只是噱头而已。增材制造已经发展很久了，鸿海 30 年前就在用增材制造技术。不看好的原因是，此项技术无法用于大量生产，不具有商业价值。

富士康为苹果代工生产 iPhone 手机已经多年。郭台铭以 3D 打印制造的手机为例，说明 3D 打印的产品只能看不能用。他说 3D 打印目前还不能加入电子元件，导致无法对电子产品进行量产，且一摔就碎。

制造业大佬看衰 3D 打印其实还有着更深层的心理原因。作为第二次工业革命时代以来“成功的进化者”之一，郭台铭带领着鸿海及富士康建立起了像恐龙一样庞大的“制造业帝国”，虽然利润微薄，但也可以凭借着“大规模流水线生产”以量取胜。可是，当侏罗纪已经过去，产业环境正在开始发生变化，白垩纪的恐龙看到以“3D 打印”为代表的哺乳动物露出小脑袋时，让它们微笑面对确实有点困难！不过，历史的真相是，哺乳动物并没有去剿灭恐龙！当然，恐龙更没有机会去剿灭哺乳动物！恐龙最终离开这个蓝色星球完全是因为自身无法适应新环境而造成的。适者生存，关于“产业进化”这一严肃命题现在已经正式摆在我们每一个人面前了！



参考：也许在苹果公司看来，3D 打印不会是无用的。手机之所以能迅速投入市场，很大的原因在于它们的零部件在设计阶段都用到过 3D 打印技术。打开 iPhone 手机就会发现，里面有不计其数的细小部件。之所以用 3D 打印，就是为了确保每个部分都能完美耦合，不会装不到一起去。郭台铭的富士康位于产业链的底端，而苹果面向产业链上游进行设计和策划。

其实，不仅仅是郭台铭，3D 打印技术的应用从诞生之时，就饱受质疑。一直以来的质疑集中在：1) 技术的可实现性：生产工艺是否可以达到产品需要的质量标准，2) 技术的经济性：大批量生产成本过高。

1.3.2 关于“3D 打印技术的可实现性”释疑

3D 打印技术的发展和完善贯穿于产业链的各个环节，具体来说主要体现在 3 个方面。

应用材料领域的不断拓宽

应用材料的发展无疑对 3D 打印技术使用领域的拓宽具有最直接的影响。20 世纪 90 年代初，塑料作为 3D 打印材料的推广应用，随之带来 3D 打印行业的第一个迅速发展阶段。21 世纪初期，随着激光烧结，特别是 DMLS（直接金属激光烧结）技术的发展，金属材料成为 3D 打印材料应用上另一个重要的突破。由此，3D 打印技术逐步进入工业部件和工具制造领域，成就了行业 2004 年之后的迅速发展。

制造精度的提高

制造精度的提高源于成型工艺的发展和数控精度的提高。如 2000 年，Objet 改进了传统的 SL 技术，制造精度从最初的毫米级提升到 20mm 以下。而现阶段，主流的 3D 打印工艺的制造精度较早期已经大大提高，可以满足大多数制造工业的精度要求。

近些年来，随着一些技术路径已经基本成形，工艺原理上的突破已经较少。技术更新更多地集中于数控系统精度的提高、软件的增强升级和制造速度的提升上。如 2010 年，Materialise 公司发布新的用于支持金属材料成型的软件系统。

成型技术功能性方面的拓展

3D 打印技术的发展除了在核心成型技术上的突破之外，一些衍生工艺也在不断完善。如 ZCorp（已被 3D Systems 收购）、Objet（已被 Stratasys 收购）开发的基于多种材料的混合打印技术，3D Systems、Stratasys 开发的多余材料自动清理和循环使用的工艺。在很大程度上，这些工艺在不断提升 3D 打印的整体技术性能，扩大制造的适用范围。

1.3.3 关于“3D 打印技术的经济性”释疑

就技术的经济性而言，被市场质疑较多的还在于 3D 打印较高的直接制造成本。从现阶段 3D 打印制造的成本构成来看，设备和材料占据主体部分，但两者随着技术的发展和市场规模的扩大，都存在较大的下降空间，有望带动直接制造成本的下降。

就制造速度而言，以激光烧结为例，最初的加工速度大致是 6 ccm/h (cubic centimeters per hour)，现阶段平均制造速度在 11~12 ccm/h。而基于超声波焊接 (UAM) 工艺的 3D 打印速度可以达到 492 ccm/h，制造速度的提高带来单位部件制造成本的降低。



提示：在第 2 章中将介绍其他金属成型工艺，其中：目前 SLS 的成型效率为 5~30 ccm/h (国内)、5~80 ccm/h (国外)；SLM 的成型效率为 2~10 ccm/h (国内)、2~26 ccm/h (国外)；LENS 的成型效率为 2~7 ccm/h (国内)、2~24 ccm/h (国外)；EBSM 的成型效率为 2~10 ccm/h (国内)、2~20 ccm/h (国外)。

此外，另一个直接制造成本的重要组成部分来源于打印材料。现阶段，3D 打印材料的主要类型是塑料、液态树脂和金属粉末。3D Systems 和 Stratasys 公司主要供应的是前两种材料。从公司材料的销售毛利率来看，近些年来基本保持在 50% 以上，就利润空间而言，还有较大的下降空间。此外对于金属材料，如 3D 打印领域使用最为广泛的钛、铝和不锈钢，仔细调研可以发现，目前厂商的利润空间也是非常大的。因此长期来看，价格也存在较大的下降空间。

尽管在直接制造成本上，3D 打印相对于传统制造工艺很难具备规模生产的经济性，但由于 3D 打印技术在产品制造上较高的自由度，产品设计的优化可以提升产品使用过程中的经济性。就逻辑而言，如果一个产品设计改变给相关的产业链带来的价值增值超过了成本的增加，那么意味着这种设计就更具有经济性。

如图 1-25 所示，EOS 公司根据客户需求改进了一个用于加工塑料杯子的模具设计 (模具在 **注塑成型、受压成型、冲压成型** 中必不可少)。传统设计的模具随着工作时间加长，模具表面温度不断升高，会影响塑料的冷却成型效率，导致成品率下滑。使用 3D 打印技术改进后的模具设计，在内部加设了用于降低模具温度的环状导管。采用这一设计的模具在使用中提高制造速度约 40%，同时还降低了能源消耗。每年节省的制造成本约为 24 000 欧元，带来的经济价值明显高于成本增加。



图 1-25 左：使用 DMLS 技术制造的注塑模具；右：模具内部结构剖面示意图 (图片来源：EOS)

使用 3D 打印技术与传统工艺在设计理念上的一个明显差异在于：前者更适合使用一体化设计思路，将之前需要多个零部件组装的产品，转变为一次成型的制造模式。如诺斯罗普格鲁曼公司设计的用于航空环境控制系统的零部件，传统设计需要 9 个零部件组装成型，而利用 3D 打印可一体化设计并一次成型，省去了组装过程，而且从理论上来讲，零部件数目越少，产品就越安全。此外，3D 打印带来的成本节约还体现在免去了：组装线的固定资产投资，组装线的人工和能源消耗，不同组装件包装、标签、运输和库存管理成本。因此，在一些领域使用 3D 打印制造带来的经济增值可能会远大于成本。

3D 打印技术对零部件的修复也独树一帜。航空航天零件结构复杂、成本高昂，一旦出现瑕疵或缺损，只能整体更换，可能造成数十万、上百万元损失。而通过激光成型修复（Laser Forming Repair, LFR）的 3D 打印技术，如图 1-26 所示，可以用同一材料将缺损部位修补成完整形状，修复后的性能不受影响，大大节约了时间和成本。

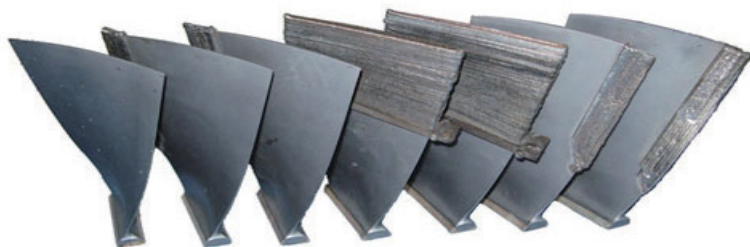


图 1-26 激光成型修复的不同缺陷形式的航空发动机叶片（图片来源：西安铂力特）

此外，3D 打印技术对产业链另一个成本节约的形式具体表现为：能源成本的节约。这一点主要体现在交通运输设备领域，部件设计的轻质化带来的燃料消耗和废气排放的减少。以航空业为例，大多数零部件的重新设计，有望降低部件重量的潜在空间达到 70%。而每架飞机减少 1kg 的重量，意味着每年的燃油成本节约在 3 000 美元左右。

3D 打印技术是典型的颠覆性技术，一台打印机可以“万能地”制造种类繁多的定制化产品，有时甚至直接打印成型而无须组装。而传统制造方式需要改变或裁剪流水线才能完成定制生产，其过程需要昂贵的设备投资和长时间的工厂停机。

1.3.4 关于“3D 打印产业的成长性”释疑

随着“个人智造”的兴起，在个人消费领域，3D 打印行业预计仍会保持相对较高的增速，有助于拉动个人使用的桌面级 3D 打印设备的需求，同时也会促进上游打印材料（主要以光敏树脂和塑料为主）的消费。

在工业消费领域，由于 3D 打印金属材料的不断发展，以及金属本身在工业制造中的广泛应用，预计以激光金属烧结为主要成型技术的 3D 打印设备，将会在未来工业领域的应用中获得相对较快的发展。中短期内，这一领域的应用仍会集中在产品设计和工具制造环节。

综合以上特点趋势，从行业发展的角度来看，整个 3D 打印产业链都存在巨大的潜在发展空间。就未来的中期需求增长而言，相对看好上游打印材料和个人 3D 打印设备的制造企业。因为

就前者而言，在通用化的技术标准不断推广的基础上，专业化的材料供应企业的发展是大势所趋。从个人消费到工业制造，无论是哪个领域引来的快速增长，对于耗材的需求都必不可少。而从长期来看，为 3D 打印量身打造数字化建模软件的应用平台公司将成为行业巨头（正如目前的微软、谷歌等应用平台公司相比于戴尔、联想、惠普等设备提供商），利用互联网思维对传统快速成型行业进行数字化、智能化的深度改造和升级，是 3D 打印行业不可阻挡的历史大势。

通常，技术的发展往往遵循一个可预期的模式，即先是萌芽，然后炒作，而后幻灭，接着才是技术成熟后的稳步爬升，最后到达应用高峰，这样构成了一个完整的技术炒作周期（Hype Cycle）。研究分析机构 Gartner 推出了 2014 年技术发展趋势报告，该报告评估了 119 个门类、2000 多种技术的成熟度，将这些技术归类到不同的炒作周期阶段。如图 1-27 所示，可以看出消费级 3D 打印（Consumer 3D Printing）目前正处于炒作期，而工业级 3D 打印（Enterprise 3D Printing）正处于技术成熟后的稳步爬升。

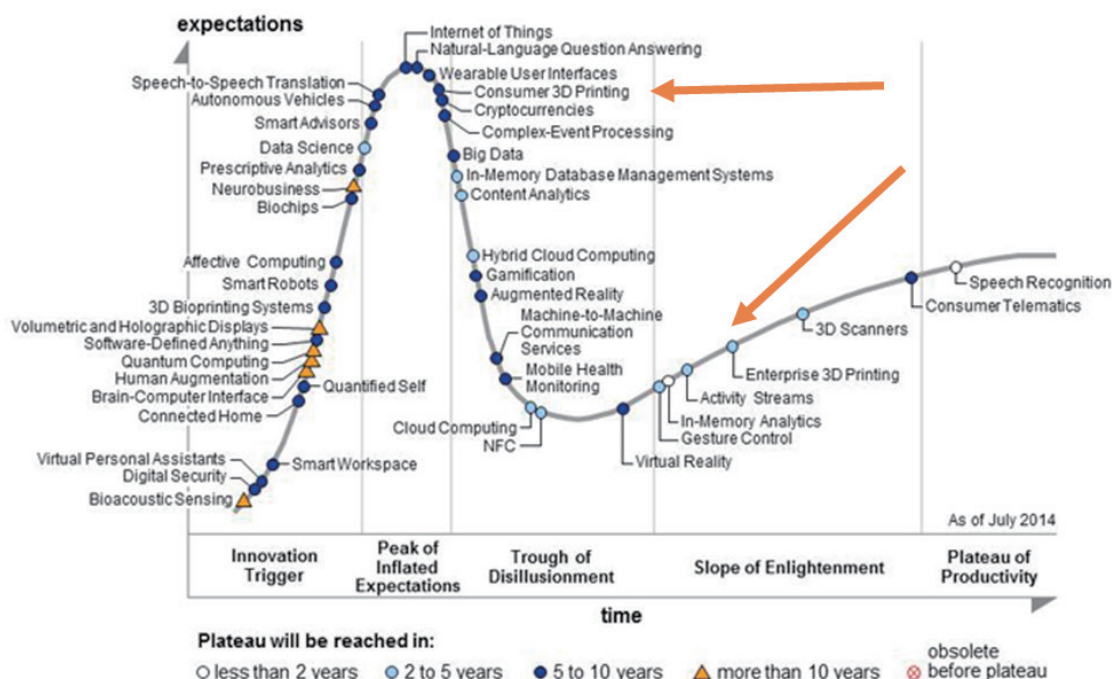


图 1-27 研究分析机构 Gartner 推出的 2014 年技术发展趋势报告

忽略 3D 打印的影响力就等于无视即将发生的颠覆，就像当年小型机厂商无视个人电脑的出现，而具有讽刺意味的是小型机“今安在”？所有颠覆性技术最初往往逊色于当时占主导地位的技术，但会不断发展，以低成本满足较高端市场的需要，然后夺取天下。“所有重要的科技都是在短时间内被过度炒热，其功能性也被高估，但从长期来看，他们造成的影响却远被低估。”美国《连线》杂志前主编安德森认为。

一个耐人寻味的现象是，制造业大佬的看衰并未影响当日 A 股市场中 3D 打印概念股的走势，概念股悉数飘红。全球 3D 打印产业已然实现了 3 年 25% 以上的复合增速，且成长还在延续。

1.4 3D智能数字化与3D打印：用“虚拟”再造“现实”

目前，全球正在兴起新一轮数字化、智能化制造浪潮。欧美等发达国家面对近年来制造业竞争力的下降，抓住以网络化为驱动的“创客运动”的发展机遇，大力倡导“再工业化、再制造化”战略。以智能数字化为核心的“第三次工业革命”引发的前提和基础是模式识别、视觉计算、自动化控制、机器学习、大规模数据挖掘等学科的成熟以及大批量低成本传感设备的普及。这种深层次的产业革命，不仅将席卷人类的体力劳动岗位，也将毫不留情地占据人类之前引以为豪的脑力劳动岗位。任何能够提取出统计规律、特征描述或编码索引的日常工作都将被自动化。可以确信的是，将来一个人薪酬的高低，将取决于他掌握智能数字化的专业程度。

而作为“第三次工业革命”的前沿代表技术——3D数字化打印，成功地将虚拟的数字智能化技术与实实在在的工业产品桥接在一起。据预测，3D打印行业的产值将在2016年达到31亿美元。作为快速成型技术的一种，3D打印以经过智能化处理后的3D数字模型文件为基础，运用粉末状金属或塑料等可黏合材料，通过逐层打印、叠加成型的方式来增量构造物体。

3D打印实质上反映了制造业向智能化不断演进的历程。由于3D打印个性化定制的特点，决定了其不具备规模经济（即产品千篇一律的“大规模生产”），相应地，3D打印技术推动的未来商业模式之一将是云制造，其由数百万个小规模、自动组织的生产节点组成。这个由众多小型制造业企业组成的超大规模分布式网络，结合云端的智能计算服务，将形成一种全新的“大规模定制”解决方案。

1.4.1 3D智能数字化设计技术的发展现状

在传统的2D打印机时代，我们可以把经计算机处理过的Word等格式的电子文档在纸张上打印出来。2D打印机对我们绝大多数人而言，仅仅是一种工具而已，因此，我们可能不太会去在意使用的是何种品牌的2D打印机，只要它能把“电子文档”转变成“纸制文档”就行。我们几乎把所有的关注点都放在了电子文档内容的设计和制作上，因为它才是我们工作的价值体现，这种情况在3D打印时代也完全一样。

尽管说到3D打印，就会让人联想到打印出来的硬件和物品，但是3D打印的神奇之处来自软件。因此，智能数字化软件是3D打印的核心，其利用计算机来生成数字化的3D模型，以便输出到3D打印机。正所谓“巧妇难为无米之炊”，缺少数字化文件支持的3D打印机将只是一个空壳。3D智能数字化设计使得消费者和制造商之间关系越来越密切，可以在产品打印出来之前对方案进行反复修改，并使用户对于定制的期望变得更强烈。

为了让计算机知道如何更好地设计形状，目前有两大类方法可以进行3D数字化。本节介绍第一大类方法，即使用智能数字化设计软件，由设计师从无到有地设计3D数字化产品（详细介绍请阅读第4章“3D智能数字化：3D打印的孪生兄弟”）。

最简单的几何表示是采用传统的建模工具，如使用SolidWorks、AutoCAD、3DS Max、Maya、Rhino3D、ZBrush等常见3D商业设计软件，还有Blender、Tinkercad、3DTin、SketchUp等多款各有特色的免费设计软件，来表达曲面网格形状。

其次是使用参数设计软件，如简单指定长、宽、高这3个参数，就能获得一个定制的茶杯

形状模型。更加智能化的是程式设计，计算机把形状的设计过程描述成一系列有特定顺序的操作步骤，有点像按照食谱而不是最终的外观来制作蛋糕。程式智能设计可以轻易地在这个蛋糕上绘制几百万个规则的精美图案，而这对于手工设计来说犹如噩梦。

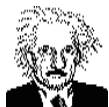
为了生成更加丰富的个性图案，还可以采用复杂的生长式智能系统，按照一套既定的生长规则加上随机扰动，随着时间的推移发展，将一颗种子形状最终生长成独一无二的定制形状。

智能化达到一定层次后，更可让设计的形状根据未知环境实时调整，适应各种物理和美学约束条件。比如，基于算法的智能设计软件能够根据物理环境调整建筑结构的形状，从而使建筑结构更稳定。

采用人工智能进行设计的另一个途径是增强人和计算机之间的交互性，用户不需要了解计算机设计的内部原理，只需从计算机推荐的参考形状中不断地做出挑选，计算机根据反馈对参考形状进行优化调整，如此反复，直到最终生成一个满意的设计。

对于要求具有复杂的内部中空、凹陷、互锁或者有大量规则细节图案的形状加工，3D 打印机是首选或唯一的制造设备。智能化设计可对零件进行优化，减轻重量，同时保持原有强度和其他关键性能；还可使产品成为一个整体，这样也减少了零部件的装配。

注意：3D 打印机往往并不能直接打印任意复杂的形状。大多数设计文件，特别是那些复杂物体的设计文件，都需要专业人员进行调整优化。此外，一名好的设计师必须考虑支撑结构，以便在 3D 打印过程中帮助物体保持形状。还有一个困难的问题是如何解决多种材料的混合制造。只有实现了混合材料打印，多元结构的部件才能一次制造出来，以避免传统的首先制造单个（不同材料）零件再组装在一起的弊端。



1.4.2 智能数字化扫描技术的发展现状

当然，并非人人都有能力自己设计 3D 形状，因此第二大类的 3D 数字化就是 3D 扫描（俗称 3D 照相），基于计算机视觉、计算机图形学、模式识别与智能系统、光机电一体化控制等技术对现实存在的 3D 物体进行扫描采集，以获得逼真的数字化重建（详细介绍请阅读第 5 章“3D 智能数字化与 3D 照相馆：科学与艺术的结合”和第 6 章“视觉计算：构建 3D 打印的杀手级应用”）。3D 扫描技术可分主动式（Active）扫描与被动式（Passive）扫描两种。

主动式扫描是对被测物体附加投射光，包括激光、可见白光、红外光、超声波与 X 射线等。其中激光线式的扫描（如手持式激光，Handhold Laser），可以扫描大型物体，但是由于每次只能投射一条光线，所以扫描速度慢。另外，由于激光会对生物体以及比较珍贵的物品造成伤害，所以不能应用于某些特定领域。而目前最新的基于结构白光（Structured Lighting）的扫描设备，能同时测量物体的一个面，点云密度大、精度高，在快速采集物体三维表面信息方面具有独特优势。除此之外还有基于时差测距（Time-of-Flight）、三角测距（Triangulation）、调变光（Modulated Lighting）和光照编码（Light Coding，如 Microsoft Kinect 设备就是采用此原理，具有实时性的特点）的主动式扫描技术等。

被动式扫描对被测物体不发射任何光，而是采集被测物表面对环境光线的反射，因不需要特殊规格的硬件，往往只需要一台或几台照相机获取多个视角的图片即可，因此成本非常便宜。

被动式重建方法，如 Autodesk 的 123D Catch，通常基于计算机三维视觉的理论方法，如立体视觉法（Stereoscopic）、从明暗恢复形状方法（Shape from Shading）、立体光度法（Photometric Stereo）和轮廓法等。被动式扫描的精度和鲁棒性受环境光照和照片质量的影响较大。

在获得 3D 扫描的原始数据后，往往还需要对其进行复杂的后处理，如将多个视角的形状片段进行对齐（Alignment）和拼接配准（Registration，也被译为注册），以统一在同一个世界坐标系下。此外还需要进行漏洞修补、噪声去除、三角化、重网格化等操作，以生成最终的高质量水密（Watertight）流形曲面。目前，还没有一种成熟的 3D 数字化技术能够对自然界的任意形状进行全自动地真实重建，如对于人体的头发等还不能获得理想的结果。因此在实际操作过程中，往往需要同时结合多种扫描技术，以及一定的手工编辑，以获得一个好的重建质量。

值得指出的是，在获得数字化模型之后，通常还需要进行个性化编辑定制，才会最终输出到 3D 打印机。这种追求高附加值的个性化定制，之前都是以较大的手工工作量为代价的，尤其是当需要“大批量定制”时。因此，为提高定制效率，智能数字化技术将发挥关键作用。比如，需要为一万名用户打印定制个性化的眼镜、服装、帽子、鞋子，如果使用人工逐一为每位用户进行手工测量和手工设计，工作量和成本都将变得不可接受。而应用智能数字化技术，如采用视觉计算方法，利用摄像头自动采集、分析提取每位用户的体貌个性特征，并自动根据视觉美感进行形状设计、颜色肤色搭配等，可极大地缩减定制周期。

可以说，**数字化是“第三次工业革命”的媒介和载体，而智能化则是手段和核心**。目前，智能化技术的应用研究尚处于起步阶段，离工业化的实际应用还有一定的距离，但最近几年发展很快。

1.4.3 智能云网：云端智能服务和云制造

通过上面的介绍可以看出，智能数字化技术涉及视觉计算、模式识别与智能系统、复杂系统与自动控制、数据挖掘与机器学习等众多“高科技”学科，普通技术人员掌握该技术的门槛很高。因此，这些技术将来会以云端智能化服务的形式提供给普通用户和开发者。以定制一双鞋子为例，普通用户只需在手机上下载一个 App 应用，给自己的双脚拍几张照片，并指定喜欢的款式和颜色，之后位于云端的智能计算服务将根据用户上传的照片重建出 3D 脚型，然后把鞋子设计出来。所涉及的复杂智能计算全都在云端完成，App 的开发者根本无须了解。用户提交订单后，系统在云制造集群中搜索到邻近的打印节点，以便快速送货上门。

以上涉及云制造的概念，其对 3D 打印这种“规模定制”的运营模式尤其关键。维基百科对云制造的定义是：“具有各种制造资源和能力，可以智能检测并连接更广泛的互联网，具备自动管理和控制能力”。每个单独的制造节点都是自主的、通过网络互联的。云制造的优点是资源可以扩展，还可自动平衡负载。制造商可以根据项目的特别需求，如本批次是定制一千件还是一万件，来构建一个临时的集群。每个云制造商的产能可能很小，但集群后的整体产能完全可以满足项目需求，且非常经济、灵活。

在本书中，我们把 3D 打印产业模式所依赖的云端智能化服务和云制造统称为“**智能云网**”（ICN, Intelligent Cloud Network），如图 1-28 所示。**智能化、云端化、网络化、数字化**将是 3D 打印未来的重要特点。



图 1-28 “智能云网”模式下的 3D 打印产业链

1.4.4 3D 打印技术的发展现状

近年来，我们从各类媒体上获得的关于 3D 打印的新闻逐渐增多，比如时尚的衣服、合脚的鞋子、营养的食物、后现代的房屋和自行车、汽车、无人飞机等都被打印出来了，3D 打印正在以一种不可思议的速度渗透进我们生活中“衣食住行”的各个方面。

3D 打印诞生于 20 世纪 80 年代，用于将虚拟世界中任意复杂的 3D 数字化模型变成客观世界中真实存在的 3D 实体。通俗来讲，只要你能设计出来，你就能够通过 3D 打印技术打印出任何你想要的个性化产品。3D 打印无须机械加工或任何模具，就可加工任意复杂的中空形状，解决了许多过去难以制造的复杂结构零件（如复杂的航空发动机叶片）的成型问题。而且产品结构越复杂，制造效率优势（研制周期缩短、原材料节省）越显著。目前 3D 打印在电影制作、游戏动漫、医疗、教育、建筑、文物考古、生产制造业都发挥了其独特的作用。

与 2D 打印机类似，3D 打印机也是由控制组件、机械组件、打印头、耗材和介质等架构组成的。3D 打印一般采用分层加工、叠加成型的方式来完成 3D 实体的打印工作。以喷墨沉积(3DP) 技术为例：每一层的打印过程分为两步，首先喷洒一层均匀的粉末，然后在需要成型的区域喷洒一层特殊胶水，胶水液滴本身很小且不易扩散，粉末遇到胶水会迅速固化黏结，而没有胶水的区域仍保持松散状态。这样在一层粉末一层胶水的交替下，实体模型将会被“打印”成型，打印完毕后只要从松散的粉末堆中取出模型即可，而剩余粉末还可循环利用。

3D 打印成为近年来的新闻热点，与 2008 年英国 RepRap 开源桌面级 3D 打印机的发布不无关系。RepRap 是 3D 桌面级打印发展的基石，直接催生了包括 MakerBot 在内的一大批廉价普及型 3D 打印机，价格从几千到几万元人民币不等。而在高精度大尺寸工业打印领域，成立已近 30 年的美国 3D Systems 和 Stratasys 两大公司占据了大部分的市场份额。当然，在这个新技术

竞争激烈的领域不乏挑战者，如 Mcor 公司 2012 年新推出的 Iris 全彩打印机只需普通 A4 办公纸作为原材料，具有超低的成本优势和绿色环保优势。在国内，由亚洲制造业协会联合华中科技大学、北京航空航天大学、清华大学等科研机构和企业共同发起的中国 3D 打印技术产业联盟于 2012 年成立。

与我们日常使用的 2D 打印机相比，3D 打印机所能使用的材料已经不再局限于墨粉和纸张。目前，3D 打印机已经能够使用各式各样的新材料（液体、粉末、塑料丝、金属、沙子、纸张，甚至巧克力、人体干细胞等），通过喷墨沉积、熔融沉积、激光烧结、立体光刻、电子束熔融、超声波固结等工艺将三维数字模型变成实物，从玩具、工具到厨房用品、建筑、时尚衣服应有尽有，甚至还可直接打印具备触感的人造耳朵、人体骨骼、人造假牙、鲜肉，以及枪支、跑车、无人飞机等。因此，如果说 2D 打印机属于一种必不可少的办公用品，那么 3D 打印机则将会成为一种使用广泛的个性化制造工具。利用 3D 打印机，未来甚至能够打印出人类（还记得第一张漫画图吗？其实科技一直在努力！）。

3D 打印技术目前面临着以下几个主要问题亟待解决。

- 一是与传统切削加工技术相比，产品尺寸精度和表面质量相差较大（制造精度一般仅相当于铸型），产品性能还达不到许多高端金属结构件的要求。
- 二是加工速度，以及大批量生产效率还比较低，不能完全满足工业领域的需求。
- 三是设备和耗材成本仍然很高，如基于金属粉末的打印成本远高于传统制造。

由此可见，3D 打印技术虽然是对传统制造技术的一次革命性突破，但它却不可能完全取代切削、铸锻等传统制造技术，两者之间应是一种相互支持与补充，共同完善与发展的良性合作关系。

不过，可喜的是，研究人员正在打破打印尺寸、材料整合、打印速度的局限，甚至还有系统正在研究将 3D 打印（添加式制造过程）和传统削减式生产过程的优势（如数控加工）相整合。这些整合方式同时使用 3D 打印和机械加工，省去后期处理过程。例如，大多数 3D 打印制造的金属部件需要人工干预进行精加工或抛光。然而，日本重型机械制造公司——松浦机械制作公司，开发了一个系统，整合 3D 打印（如激光烧结技术）和高速铣削，对打印成品的边缘进行五层增量铣削。

更重要的是，通过智能感知设备，3D 打印机还可控制制造的行为，对打印的过程进行实时监控，如产品的质量和强度，然后根据反馈信息随时做出调整，以实现闭环控制。也就是说，这台 3D 打印机具有学习和控制的能力。可以想象的是，会有专为糖尿病患者推出的食品打印机，通过微型皮肤植入物监测病人的血糖，依据每日不同的身体状况为其量身打印食物。在将来，通过把人工智能从计算机拓展到现实世界，还可打印具备感知和学习能力的智能物品。此时，3D 打印机就是新一代智能机器人，它们能设计、制造、修理、回收其他机器，甚至能够改进和升级机器自身，达到“机器制造机器”的新境界。

3D 智能数字化与 3D 打印技术相结合所带来的优势，不仅仅在于通过复制手段真实还原现实世界，而且还可以在 3D 数字化的基础之上，通过再设计工作，创造出一个更加美好的世界来。以电影《阿凡达》为例，很多美轮美奂的场景都无法从现实中直接拍摄，而通过数字化的艺术设计，

再使用 3D 打印机直接打印出来，这样不仅免去了费时费力的手工制作，而且获得了超越现实的逼真效果。3D 智能数字化与 3D 打印的完美结合，将实现用“虚拟”再造“现实”的崭新境界。

1.5 创客DIY：新工业革命的启蒙运动

“创客”（Maker）指喜欢动手制作，努力把各种创意转变为现实产品的人。他们会使用 3D 打印机、**数控机床**（Computer Numerical Control, **CNC**）、电子电路、激光切割机、3D 智能数字化技术等功能强大的数字桌面工具进行创造。创客，既是工具的发明者，也是工具的使用者。世界，正在因为他们而改变。

实际上，3D 打印技术 30 年前就已诞生，但并不广为人知，而被冠以“增材制造”、“快速成型”等专业术语，且一般都只应用在工业制造领域和大型实验室里。直到 2008 年，英国一名叫 Adrian Bowyer 的创客发布了第一款开源的桌面级 3D 打印机 RepRap，并把机械设计图纸、电路图纸、控制源代码无偿地放到网上供人免费下载。几年下来，便使得原本极其昂贵（几十万元起价）的 3D 打印机变成现在几千元即可买到，走入了普通家庭。家用或商务 3D 打印的普及正是制造业大众化的重要标志。实际上，近两年开始**使用“3D 打印”这个名称，而不是“增材制造”这个术语，就是大众化的直接体现。**

3D 打印让新产品的设计和测试更加快捷，现在有了一个想法，创客立即就可以运用数字化工具和桌面级 3D 打印机实现，看它到底能不能用、好不好用。产品研发周期缩短了许多，实在非常方便。等到正式生产时，完全可以把批量打印的任务外包给专业的在线 3D 打印公司，他们可以提供各种精细的工业材料以供选择（包括金属、塑料和玻璃等）。

3D 打印将越来越成为 **DIY（Do It Yourself，自己动手做）** 制作过程的核心工具，与之紧密相关的其他元素也在蓬勃发展，包括免费或低成本 3D 建模和扫描工具（用于设计阶段）、分享网站（用于营销及配送阶段）、投资网站（用于资金筹集阶段），以及新的开放式设计理念（行业合作）。所有这些发展促成了几乎人人都能成为制造者，或者为制造过程做贡献，制造者与消费者之间的界限将会变得越来越模糊。对于 3D 打印设备使用技术和相关 3D 软件的教学推广也在逐渐展开，许多高校和研究机构已经开设了相关课程和实验室。会有更多的消费者像使用个人电脑一样，自如地掌握 3D 打印机的操作，潜在的消费需求会不断爆发。

创客崇尚为兴趣爱好去做、崇尚个性化定制、崇尚开源共享。实际上，开源共享体现的是一种自信，相信即使把所有技术细节都免费公布了，世界上也没人能在这个小领域做得比自己更好！现在，不仅有开源软件、开源硬件，在互联网上还有很多的开源社区，聚集了众多志同道合者，创客们使用数字桌面工具捕捉创意，在开源社区分享设计成果，一起来享受 DIY 的乐趣。

在 21 世纪，我们每个人都希望使用与众不同的产品，正如我们每个人都与众不同一样。长尾的利基市场上充满着无穷的商机，个性化定制的产品往往利润丰厚。创客们正在形成一种新的工业组织模式，以兴趣为驱动，以项目为导向，公司规模更小，趋于虚拟、非正式，他们在运营中组队与重组。如果项目足够有趣，就会吸引顶级人才纷至沓来。这些创客公司正在上演一出“以小博大”的拿手好戏：团队成员远远少于传统的大公司，但创新能力却高于大公司。

创客们不仅仅在细分市场上表现得游刃有余，而且在大众市场上也频频掀起翻天巨浪，一次又一次地改写着人类文明的历史进程。

1.5.1 以小博大：创客挑战巨头公司

Local Motors 是一家基于网站社区平台建立的汽车制造企业，主要面向个性化的小众汽车市场。Local Motors 成功开发并销售的 Rally Fighter 车型，售价 75 000 美元。开发时间仅 18 个月，开发成本 300 万美元；而通用汽车开发沃特使用 6 年时间，耗费 65 亿美元。2011 年，美国国防高级研究计划局（DARPA）向公众收集灵感，为标志性的军用悍马车设计替代品，而 Local Motors 在短短 14 周内——约为汽车业平均制造时间的 1/5，就将获奖设计打造成了可运行的原型，如图 1-29 所示，彰显出创客军团惊人的实力和热情。



图 1-29 这辆车可能会取代军用悍马，由创客们在短短 14 周内打造，为汽车业平均制造时间的 1/5
(图片来源：Local Motors)

2013 年，鹅卵石手表（Pebble Watch）上市，这个团队当初只有 4 个人，而他们现在已经打败了索尼的智能手表 SmartWatch。鹅卵石手表是个 Kickstarter 的投资项目，如果 4 个人在尖端科技手表上可以打败索尼，这 4 个人很有可能创造下一个索尼。他们是从创客运动起家的标准创业家，从事的是个人制造业，他们有自己的优势：开发更快、更灵活，善用网络，且不受官僚体制拖累。

让我们再回溯反思一下。假使没有乔布斯，很难想象苹果会成长为全球市值最高的公司。乔教主出走，苹果衰弱，衰弱到几乎破产；乔教主回归，苹果兴盛，首先拿 iPod MP3 直接让索尼雄霸世界多年的随身听彻底绝迹，接着用 iPhone 手机直接让世界排名第一的手机公司诺基亚宣布不再做手机。要知道，苹果只是负责设计手机，自己从来就没有生产过手机，都外包给郭台铭的富士康公司去制造。因此可以说，苹果几乎是赤手空拳就把诺基亚打败了。因此在新工业革命时代，借助“取之不尽、用之不竭”的网络知识共享和协作，个人创造的威力将被放大很多倍，有时一个创客就足以推翻一个传统的产业帝国，可谓“百万军中取上将首级，如探囊取物耳”。

我们再举一些例子。之前谁又能相信，Google 的创始人，两位年轻的斯坦福大学肄业博士生仅凭软件算法就挑战掉了曾如日中天的雅虎帝国呢？之后，在社交网络领域，一位 80 后哈佛大学肄业生一个人打造的 Facebook，又让谷歌这个巨无霸难望其项背。同样，在中国，马化腾的 QQ 硬是把微软的 MSN 挤出了国门。

Facebook 一开始是从地下室起家的，在互联网时代开公司非常容易，你只需要一台计算机。而现在数字化制造业几乎也一样容易。没有任何观念表明，创客运动不会产生一家数十亿的公司。我们已经看到了至少几十家由创客运动起家的千万美元公司，还有几家是几亿美元的公司，而从几个亿变成几十个亿并不难。

最后，我们再介绍美国一位名叫埃隆·马斯克（Elon Musk）的创客。他被认为是电影《钢铁侠》的角色原型。2002 年，他把自己与人合伙创办的 PayPal（世界最大的网络支付平台，中国的支付宝与其类似）以 15 亿美元的价格卖给了全球最大的网上商店“eBay”。之后他创建了特斯拉（Tesla）公司，生产出世界上第一辆能在 4s 内从 0 加速到 100km 时速的电动跑车，并成功量产。2010 年，他创办的另一家公司 SpaceX 所发射的猎鹰 9 号火箭成功将“龙飞船”发射到地球轨道，这是全球有史以来首次由私人企业发射到太空并能顺利折返的飞船。整个宇航界为之震动。猎鹰 9 号火箭的标准发射费用为 5 400 万美元，而美国航空航天局（NASA）的为 4.35 亿美元。NASA 随后宣布，美国所有航天飞机 2011 年退役以后将不再新造，而是委托给像 SpaceX 这样的私人公司将物资补给送入国际空间站，并与之签署了一份价值 16 亿美元的合同。可以说，在火箭制造与发射方面，马斯克以一己之力打败了 NASA！马斯克还计划发明一种可以重复使用的火箭，并希望在 10 年内实现人类移民火星的梦想。

1.5.2 聚沙成塔：改变工业社会的组成结构

在笔者眼中，创客所引发的新工业社会将与传统的工业社会有明显的不同。在组成结构上，如图 1-30 左边所示，传统工业社会是典型的金字塔形，巨头公司占据在塔尖，通过垄断等手段攫取了绝大多数利润。而在以创客为主体的“个人智造”、“家庭智造”、“网络社区智造”时代，如图 1-30 中间所示，社会结构将是合理的橄榄球形（或称为纺锤形），大多数制造个体都能活得很滋润，因为市场已经充分细分和个性定制化，大家只要把自己所在的那“一亩三分地”做得足够专业即可。

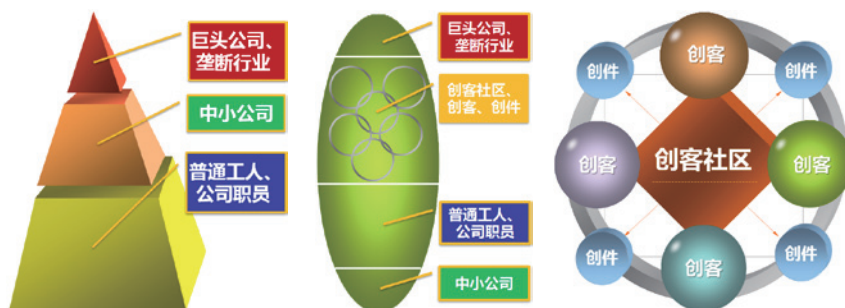


图 1-30 创客运动将引发工业社会组成结构的改变

如果你在开发一个新产品时，不了解某一相关领域的知识，没关系，如图 1-30 右边所示，你可以去**创客社区（Maker Community）**购买现成的标准技术组件（称之为**创件 Makeware**，跟软件和硬件一个道理），或者干脆在社区里雇用一个懂行的**创客（Maker）**！

当越来越多的个人创立自己的公司，不再被动地为别人打工时，自主意识、创造意识将发挥到最大，也会有越来越强烈的愿望为自己争取社会地位和权利。同时在创客社区的凝聚和团结之

下，所形成的**社会化制造**新模式将会对巨头公司和垄断行业构成有力的冲击，直到形成一种新的工业社会结构：一大部分人通过自己的努力创造过上了中产小康的生活，社会的财富均衡化。所谓**社会化制造（Social Manufacturing）**，也被称为**协作制造（Collaborative Manufacturing）**，指通过网络平台或数据挖掘手段，让用户群体提前参与（或计算分析得出用户群体的潜在需求）到产品的创意、设计、反馈和制造过程中，从而主动、及时地将可自由配置的网络化制造能力和潜在的社会需求有效连接起来，不仅创客之间可以协作共赢，而且消费者本身也因为参与进来而成为了（部分）生产者，加上制造设备（如3D打印机）也可通过租借的方式自由组合或直接将生产加工外包出去，制造于是变成了一种高度灵活的、产品可个性化定制的众包（Crowdsourcing）服务。

数字化时代正有幸迎来人类有史以来第一个“个人英雄机会”的全面爆发期，因为我们生来都是创客，从你小时候学会捏人生的第一个泥巴开始。DIY开启创新之源，兴趣在任何时候都比能力更加重要。3D打印和3D智能数字化工具正成为新一代创客手中的“双节棍”，用来创造世界的强大武器。在创客的世界里，没有永远的成功者，也不会有永远的失败者。因为“技不压身”，你总有翻身的那一天！踏着这一浪高过一浪的创新潮流，创客DIY正在发展成为新一轮工业革命的启蒙运动。

1.6 “中国制造”向“中国智造”转变的机遇

“科技史上最不可思议的事就是中国没能维持住其技术霸主地位。”工业革命史学家、美国人文与科学院院士乔尔·莫基尔如此评论。以“四大发明”为标志，古代中国一直在多个工业领域居世界领先地位。然而到了18世纪初，中国的领导地位逐渐丧失，欧洲国家以及美国成为世界工业大国。

终于，在经历了300年的沉沦之后，在2010年中国重新成为世界头号制造业大国，此前美国在这个位置上盘踞了一个多世纪。虽然重回第一，我国制造业发展却呈现出明显的不足：大而不强、徘徊于产业链低端、自主创新能力弱、资源配置能力弱、能源消耗大等。

1.6.1 “中国制造”需转型升级

据《劳动报》报道，2011年我国彩电、手机、计算机等主要电子产品产量占全球出货量的比重分别达到48.8%、70.6%和90.6%，但电子信息百强企业实现的利润仅为美国苹果公司利润的40%。当“Made in China”（中国制造）风靡全世界，我们却几乎没有掌握什么核心技术。有人曾算过这样一笔账，中国出口8亿件衬衫，才能换回一架波音747飞机。再如，一台售价499美元的第一代iPad，只有33美元是加工劳动力成本，而最终在中国装配的成本只占8美元。这深刻揭示了“**中国制造，美国利润**”现象。中国社科院工业经济研究所发布的《中国产业竞争力报告（2012）》指出，我国制造业产品在国际市场上尽管拥有较高的占有率，但仍处于国际产业链的低端加工组装环节，产业技术含量和附加值偏低。

中国凭借人力和土地资源的价格优势，仅仅用了30年时间便取得了“世界工厂”的地位。但现在，随着土地和人力资源的价格上涨，“中国制造”的廉价优势正在逐步丧失。目前，在中

国雇用一名工人的花费，在泰国可以雇用 1.5 名工人，在菲律宾可以雇用 2.5 名，印尼可以雇用 3.5 名。以制鞋、服装等为代表的劳动密集型产业正在向东南亚或非洲等生产成本更低的地区转移。因此，从形势的紧迫性上看，中国制造业必须转型升级，向制造业的中上游发展。

值得一提的是，虽然郭台铭嘴上说不看好 3D 打印，但对于“中国制造”目前面临的问题和遇到的困难，他却有着非常清醒的认识。作为全球第一大代工厂商，富士康仅在中国大陆的工人就有 160 万。2011 年年底，富士康公布了“百万机器人大军计划”（即用 3 年时间生产出 100 万台智能化工业机器人，并逐步投放到富士康的各条生产线上）。郭台铭说：“富士康的年轻人将重新学习操控机器人软件、应用和维修，变为机器人的应用工程师和软件工程师，通过操作机器人的手和关节来完成生产。把单调重复的工作交给机器人，这是中国制造业向世界发出的一个信号”。

娃哈哈集团董事长兼总经理宗庆后在接受《中国经济周刊》采访时表示：当前，中国不应再做“世界工厂”、“廉价的工厂”。东方电气集团董事长王计的观点也颇具代表性，他说，以前我们是向国外买技术，随着中国制造业不断做强做大，已很难买到甚至买不到核心技术，必须走自主创新的路子。

因此，中国制造业的当务之急是增强可持续发展能力和自主创新能力。过去 30 年我们制造业取得巨大成功依赖的是我们“勤劳的双手”，而未来 30 年要继续成为世界制造业巨头必须依靠我们“智慧的大脑”。

1.6.2 来自“德国制造”的启示

宝马、奔驰、大众、西门子，德国“牛气十足的工业制造业”让很多中国人印象深刻。《日本经济新闻》称，英国产业革命带来了世界工业大发展的契机，但是真正将世界工业制造推向高潮的是“德国制造”。从造船、钻探机械制造到高速列车、地铁、汽车、飞机等的制造，德国从来没有离开过世界前三名的位置。日本的高速铁路最初也是引进的德国技术。

你可能想象不到的是，在 125 年前，英国人对德国制造的评价却是“厚颜无耻”。相对于英法来说，德国属于后发展国家，在夹缝中追求强国梦的德国人不得不“不择手段”，仿造英、法、美等国的产品，并以廉价销售冲击市场。偷窃设计、复制产品、伪造制造厂商……德国产品因此被扣上那顶不光彩的帽子。1887 年 8 月 23 日，英国议会通过了侮辱性的商标法条款，规定所有从德国进口的产品都须注明“Made in Germany”，以此将劣质的德国货与优质的英国产品区分开来。

8 月 23 日于是成了“德国制造”的誕生日。《德国之声》称，从 125 年前的那个日子后，德国人争气地让自己销售到世界各国的产品比当地货的口碑还要好。《南德意志报》称，“德国制造”125 年的历史就像一个童话。它也是德国在二战后崛起的密码、欧债危机中仍“一枝独秀”的答案。“德国制造”已成了世界市场上“质量和信誉”的代名词。最近，德国机械及设备制造协会宣布，德国机械制造业出口继续占据世界第一的位置。

“德国制造”长盛不衰在于“质量”和“创新”。“德国制造”的成功，首先要得益于德国严格、健全的质量认证和监督体系。目前，德国最主要的标准制定机构为德国标准化协会，每年

发布上千条行业标准，其中约 90% 被欧洲及世界各国采用。这些标准织成一个密网，严格限制住企业的一举一动，从而保证了产品质量。事实上，早在 1876 年美国费城世博会上德国产品被视为廉价的劣质品时，德国学者就开始呼吁工业界清醒过来：占领全球市场靠的不是廉价产品，而是质量。20 年后，英国罗斯伯里伯爵 1896 年就表示：“德国让我感到恐惧，德国人把所有的一切……做成绝对的完美。我们超过德国了吗？刚好相反，我们落后了”。

“德国制造”的成功法宝还在于不断创新，产品不仅是在德国制造，更是在德国创造。生产之前的设计和创造才是最重要的，从新型动圈式高保真耳机，到高性能竞赛用摩托车，再到自动化的仿生物机器人和有机太阳能电池，“德国制造”以其先进的技术和舒适独特的设计，成为一张代表严谨和可靠的国家名片，得到广泛认可和推崇。

“德国制造”根植于科研机构。300 多所高等院校、数以百计的研究机构，“制造科技”都是其研究的重点。德国制造业还重视加强中小企业与研究机构联系，让中小企业参与尖端技术领域研究，促进研发和科技成果转化成为真正的生产力。最重要的是，德国还有一支将“制造科技转变成产品”的高水准技术工人队伍。而这离不开德国的技术教育。在德国人看来，职业技术教育与普通学历教育同等重要。

20 世纪 80 年代，德国和日本同样经历了货币大幅升值，两国以出口为导向的经济同样面临巨大压力。但与日本不同，德国没有紧紧盯住汇率，而是逼着企业在压力下不断挖掘潜力，自我改造，结果“德国制造”至今能够领先世界。实际上，“德国制造”的严谨和与时俱进等民族性格都是在市场竞争压力下养成的。

1.6.3 “中国智造”的发展机遇

由“智能数字化制造”引发的第三次工业革命山雨欲来。中国正处于从“中国制造”向“中国智造”迈进的重要时期，3D 智能数字化及 3D 打印技术可以让国内的设计师和工程师从产品制造工艺的束缚中解放出来，更加专注于产品本身的智力创造，大跨步进入想法到产品（Mind to Product）的“所想即所得”全新智造时代。据预测，我国的 3D 打印市场将在 3 年内从目前的约 10 亿元人民币增长到 100 亿元，中国将超越美国成为全球最大的市场。3D 智能数字化和 3D 打印的产业化无疑将为促进我国传统产业升级、彻底摆脱长期处于制造业产业链底端的尴尬局面发挥十分重要的推进作用。

“中国智造”的转型基础和步骤

从世界著名咨询公司德勤发布的《2013 全球制造业竞争力指数》可以看到，位居竞争力首位的是中国，且预计 5 年后中国仍然稳居第一。中国“世界工厂”的地位并没有被削弱，因为中国制造业仍然具有以下三大优势。

首先，中国有迅速扩大的市场。2012 年全国社会消费品零售总额超 20 万亿元，2008 年只有 10 万亿元，5 年翻了一番，这在国际上是不可思议的。并且，中国的消费市场还将继续扩大。

其次，虽然中国劳动力成本越来越高，但劳动力的素质近些年来明显提高。中国劳动力的“性价比”在国际上还是很有竞争力的。越来越多的跨国企业把研发中心搬到了中国，因为他们看到了中国更优质、更庞大的智力资源——人才！有跨国公司做过对比，在中国用一个博士，工

资成本仅为美国的 1/4。

第三，“可靠的供应链”便是中国制造业的另一大优势。现代工业制造都是分工合作的，需要大量的配套商，比如汽车产业就很典型。中国近年来成功建立起了世界上相对很有优势的产业配套体系，而包括越南在内的一些东盟国家还不具备。中国集中力量将供应链本土化并建立创新中心，从而被德勤视为唯一一个能够跟发达国家并驾齐驱，拥有同样供应商网络优势的新兴国家。

尽管“**科技革命**”有可能在欧美发达国家率先突破，但基于我国“世界工厂”的制造优势和潜力巨大的市场空间，“**工业革命**”则有望在中国大地成燎原之势，在欧、美、日陷入危机阴影自顾不暇时，我们应牢牢把握历史机遇实现“弯道超车”。

对于个人和小型企业而言，可以先在跨国公司无力顾及的利基行业（用户相对小众但利润颇高），利用本土化的优势跟国外中小公司的产品进行竞争。以 3D 人像扫描仪为例，国外的产品需要 15 ~ 20 万元，而根据作者的了解，国内的用户一般都希望价格不超过几万元。这个要求其实并不苛刻，并且仍可保证较大的利润空间。此外，国外 3D 扫描软件一般没有中文版本，让普通用户一下子很难适应，这也是国内中小公司的机会。

之所以一开始选择利基行业，就是为了避免与国外大公司硬碰硬竞争。在逐渐占领各个利基行业之后，通过强强联手，共同搭建平台，并慢慢向大众领域渗透。这样，利用本土化的优势，并借助高科技工具（如 3D 打印和 3D 智能数字化），将产业链由低端向高端转移，将产品由利基行业向大众消费行业转移，最终完成由点到线、到面、再到体的“中国智造”战略转型。

3D 打印与 3D 智能数字化助力“中国智造”

2011 年，我国十大科技进展之一是成功研制世界上最大的 3D 打印机。与传统的“切削去除材料”的加工技术（如 3D 雕刻）完全不同，3D 打印采用分层加工、叠加成型的方式“逐层增加材料”来生成 3D 实体，这既是制造工艺的原理创新，也是应用数字化技术的产品创新。中国正在迎来全新的“智”时代，最鲜明的表征就是，智能化和数字化产品正以不可阻挡之势改变着我们的生活，也影响着制造业发展的格局。

3D 打印技术的深层意义在于，它将有可能改变传统工业格局和大规模生产方式的理念，将极大地降低制造业建立工厂的基本要求和投资额度，引发新一轮的小型企业兴旺和扩张的潮流。3D 打印能有力地挑战现有垄断化、大规模化的工业模式，而以其所缺乏的灵活创新的生产机制诞生大量有竞争力的实力企业。3D 打印宣告了“厂大人多”时代的终结，尤其是那些产品更替时间短、每几个月就会发生一次产品换代的行业。3D 打印技术将加速科技创新，这让我国长期处于追赶和规模扩张型的工业产业模式看到了改变的希望，给“中国制造”向“中国智造”的转型提供了新动力。

“中国智造”一个重要方面的体现就是，要应用智能数字化技术极大地提高各种设计制造工艺的精度和效率，大幅度提升制造工艺水平。如我国自主研发的 ARJ21 飞机采用了 3D 数字化设计技术和并行工程方法，最终实现了大部件对接一次成功。另一个重要方面体现在生产过程中，比如数字化车间乃至数字化工厂，生产系统向着具有感知、决策、执行能力的智能化系统发展。走向“中国智造”，意味着企业需要坚持科技创新，逐步将 3D 打印、人工智能、机器人和数字化等高新技术服务集成应用于制造业之中，做大、做强“高端制造”。

3D 打印与 3D 智能数字化的产业化机遇

就 3D 打印技术领域本身来说，我们和国际相比虽还有一定的差距，但已不太大，大家已基本站在同一条起跑线上。国内的相关公司企业，如中航重机、南风股份、银邦股份、大族激光等，都与国际发展的趋势相接轨。我国自 20 世纪 90 年代初开始追踪 3D 打印技术研究，目前已取得了一批基础研究和产业化成果，部分甚至处于世界领先水平。例如，北京航空航天大学、西北工业大学开展的金属熔敷成型技术研究，在国际上首次突破了钛合金、超高强度钢等难加工大型复杂整体关键构件激光成型工艺，已在大型运输机、舰载机、我国第一款本土商用大型客机 C919、歼击机中装机应用。又比如，中国科学院自动化研究所的复杂系统管理与控制国家重点实验室所研制的高精度桌面级 3D 打印机（如图 1-31 所示），利用数字光处理技术实现光敏树脂的成型，层厚可达 25mm。



图 1-31 中国科学院研制的“天工 II 号”桌面级 3D 打印机，层厚可达 25mm
（图片来源：中国科学院自动化研究所）

然而，在商业化市场领域，高端 3D 打印设备和材料的定价权掌握在国外少数几家公司手中。这些高端设备售价非常昂贵，而国内尚缺乏相关的替代品，因此极大地增加了 3D 打印行业的运营成本。因此，加强我国在 3D 打印关键技术领域的研发投入，如设备和功能材料的制备、混合材料打印、智能控制问题的解决、激光器 / 喷嘴等核心元件的研制等，并进行商业化生产销售，对市面上的国外同类产品进行价格上的有效制衡，是支撑“中国智造”模式的前提和保障。

我国政府目前对此也很重视。2012 年 12 月，工业和信息化部副部长苏波表示，3D 打印将深刻影响制造业的未来，成为未来新的经济增长点。为加快推动 3D 打印制造技术的研发和产业化，我国将加强制度顶层设计和统筹规划，加大财税政策引导力度，适时筹建相关制造行业组织。这也是官方正式为 3D 打印技术在我国的发展定调。

此外，作为 3D 打印的前端和上游产业链，3D 智能数字化扫描是一项关键技术。因为对于家庭的日常 3D 打印任务而言，最重要的一个环节是进行全（半）自动的数字化建模。目前国内外的 3D 扫描设备在采集质量和速度上和国外的同类产品相差不大，价格却可仅为 1/4 左右。然而，在市场化产业化上仍有明显差距，大部分产品都出自小型公司，尚未形成有影响力的品牌。这

方面有待于政府和商业机构进一步加大支持和投入。待时机成熟，完全可以使得国产 3D 扫描设备占据绝大部分国内市场甚至国际市场。

特别需要指出的是，在 3D 数字处理软件方面，我国与国外的差距仍然较大。实际上，待 3D 硬件设备成熟之后，国际 3D 打印市场的核心竞争将转移到相关的配套软件上来。目前国内 3D 扫描厂商大多数直接采用国外的大型成熟商业软件，如美国的 Geomagic Studio 等。原因在于 3D 数字处理软件的研发需要巨额的资金投入和长期的技术积累，目前国内的小型公司难以承受研发风险，以及可能的知识产权侵权风险。但从长远来看，拥有国产化的 3D 数字处理软件是十分必要的，且是可行的，因为目前国内的科研单位（如中国科学院、浙江大学、清华大学、北京航空航天大学等）已基本解决了相关的技术难点，只是没有资金实力形成功能完整的大型软件系统。

除了大型的 3D 数字处理软件，其实还有很多可实现单一特色功能的智能化软件值得关注。前面已提到，智能数字化技术的应用目前在国内和国外都处于起步阶段，差距不大。而且这类功能单一软件的研发风险小，可作为缩小与国外整体差距的另一个突破口。以全球第一家专为 3D 打印开发 App 应用软件的毛豆科技（<http://www.moredo.cc>）为例，这是一家创立于北京的高科技公司，已开发了多款便于普通用户（特别是新手、儿童）手机平板操作的 3D 智能化建模软件，比如 3D 积木、3D 印章、3D 阴影、3D 沙漏、3D 陶艺、基于手画的 3D 检索与建模等，如图 1-32 所示。虽然每款软件功能单一，但所有软件共同形成了一个完整的 App Store 体系，并为加盟商、用户、开发者、美工、创意师打造了一个良好可持续的生态链。

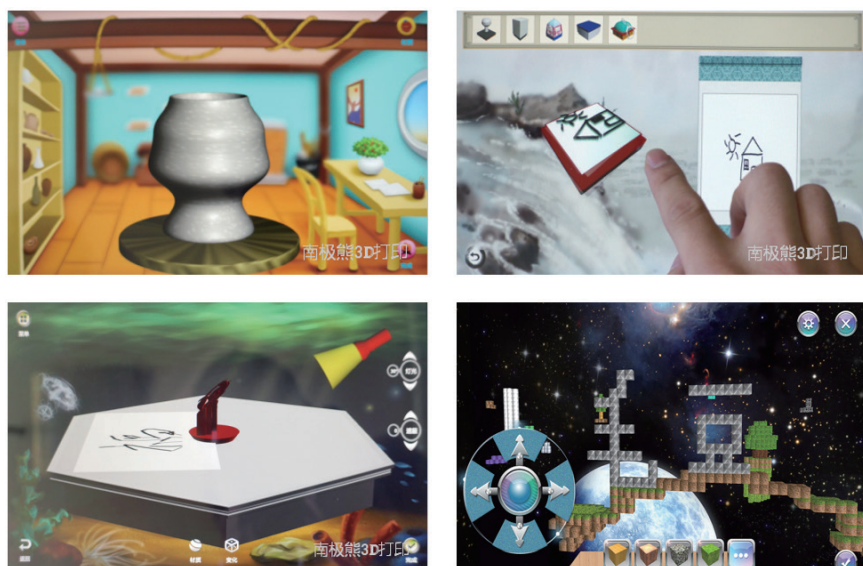


图 1-32 专为 3D 打印开发简单易用的 3D 智能化建模软件，如 3D 陶艺、3D 印章、3D 阴影、3D 积木（图片来源：毛豆科技、南极熊 3D 打印网 <http://www.dddyin.com>）

可以预见的是，以 3D 打印、视觉计算、模式识别、机器学习等为代表的智能化技术将获得广泛的工业化应用，配以低成本的传感设备，可以进行自动感知捕获、特征提取、统计分析，以及智能化定制设计，以满足高附加值“批量定制”的工业需求。

建立完善“中国智造”的产业生态圈

2010年，中国大约有1.3亿人从事制造业，约占全球制造业工人的40%。同时，中国的竞争优势早已不仅仅是低成本的普通劳动力，还有大量雇用成本适中的中等技术人才、工程师、科研人员，以及完善的产业链条、高度适应性的巨大产能和本身巨大的市场。中国的研发设计能力实际上是过剩的，这从目前市面上琳琅满目的国产手机种类以及所建立的庞大“山寨帝国”就可窥豹一斑。当前工业制造的主要门槛是复杂的制造工艺和设备，而一站式3D打印机的主要优势恰恰是零技能制造，这给我国广大技术人员研发高附加值、个性化定制的创新产品开启了广阔天地。同时，中国巨大的人口基数，又可把原本小众的利基市场变成大众市场。

时下，有人把“智造”形象地比喻为2.5产业，即介于第二产业和第三产业之间的产业，生产和服务高度一体化。也即既具备创新产品研发中心等第二产业职能，又有增值服务、贸易等第三产业职能。通过虚拟工厂为表现手法，把整个产业价值链联系起来。我国要完成向“中国智造”产业模式的转变，关键要形成一大批能够以3D产品创意设计、生产、服务为职业的群体，建立完善良性循环而非恶性竞争的创新生态圈，这方面可借鉴美国Shapeways和Quirky公司的设计、制造、销售全产业链模式。

我国目前已拥有较大规模的产品设计人员。当前存在的问题是缺乏有保障的生态环境支持这些设计人员去原创自己的风格，摆脱低水平仿造、低水平收入的恶性循环。这方面需要国家出台相关的知识产权保护法案，以及提供政策上的支持（如建立类似于Kickstarter的融资平台），还有营造创新文化氛围。3D打印机、数控机床及以Arduinos为代表的开源硬件平台降低了创新门槛，结合我国处于零部件供应链中心（如深圳华强北，如图1-33所示）的优势地位，可完成所有元器件的一站式采购，极大地提高了产品的研发速度和降低了研发成本。此外，我国还需进一步加强产业创新人才的教育和培训，整体提升国人的动手能力和DIY兴趣，可将3D打印技术纳入中学和大学的学科建设体系，增加必修环节和实训项目，为以后类似Shapeways和Quirky模式在中国的产业化形成提供相应的人才储备和技术储备。

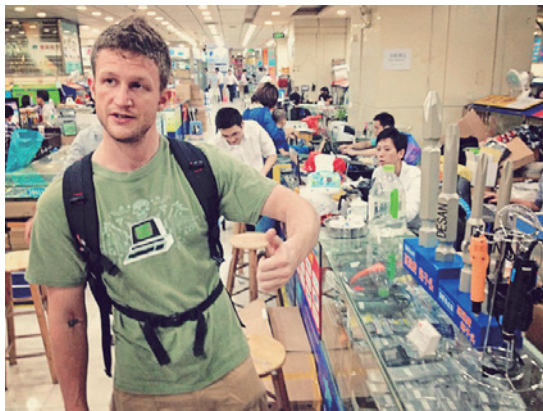


图 1-33 MakerBot 的联合创始人 Zach 选择来华创业，主要因为华强北的供应链中心地位
(图片来源：Joe Heitzeberg)

近些年来我国经济中“去实业化”现象开始急剧蔓延。以高速膨胀发展的房地产和金融市场为代表的虚拟经济，正在严重冲击我国实体经济和制造业的发展。“中国制造”发展势头放缓

已成为不争的事实。虚拟经济“一夜暴富”，也导致贫富分化差距拉大和社会心理失衡。因此，消除虚拟领域的暴利，遏制各种投机行为非常紧迫，实现社会利润率平均化，让每一位充满创造热情的年轻人有定所，以便安心创业而无后顾之忧，同时促使社会资本尽快回归制造业，为“中国制造”向“中国智造”的战略转型打下坚实的社会基础。

目前我国政府正在加大对“高端装备制造业”的支持力度，已将其列入国家七大战略性新兴产业之一，而“智能制造装备”又是其重点发展方向。《国务院关于加快培育和发展战略性新兴产业的决定》中明确指出“强化基础配套能力，积极发展以数字化、柔性化及系统集成技术为核心的智能制造装备”，并出台了相关配套政策措施。按照《智能制造装备产业“十二五”发展规划》的要求，到2015年，我国智能装备制造业的销售收入将超过1万亿元，年均增长率超过25%，工业增加值率达到35%。本土化智能制造装备的国内市场占有率将超过30%。

管理技术创新同样也是“中国智造”的重要内容。通过智能数字化系统的应用，在互联网、物联网、云计算等技术的强力支持下，使企业能对产品整个生命周期全部数据进行统一管理，建立从企业内各部门到用户之间的信息集成，从而有效地提高市场反应速度和产品开发速度。当前，十分迫切的是对我国传统制造业加快进行数字化、智能化技术改造提升。同时，智能制造技术将使制造过程的物流、信息流、资金流高效协调，大大提高制造的效率和效益，减少资源消耗，降低对环境的不良影响，成就“绿色制造”。

最后我们举一个例子说明“中国智造”的优势所在。如前所述，要打印一件3D物品，目前技术上还没有一套全自动的解决方案，仍需要大量复杂的智力和手工劳动，如3D形状的数字化扫描过程、产品创意的智能化设计、3D打印产品的清理和抛光上色等。在欧美等发达国家，人工费用和设计费用都非常昂贵，这样导致设计和打印一件3D产品价格不菲。以国外一家3D照相馆为例，如图1-34所示，其出售3D扫描和3D打印的人物雕像，一个6英寸的全彩雕像成本价约为2493元人民币，这个价位在国内几乎没有可行性。而在国内，完全可以使用国产的智能扫描设备，经设计师使用智能化软件定制加工之后，再采用低成本的单色材料，并利用低成本的单色3D打印机（如RepRap、MakerBot等）将模型打印出来，最后雇用极具价格优势的美工流水线进行人工上色，全部成本在“中国智造”模式下可控制在100元人民币以内。



图 1-34 国外一个6英寸的全彩雕像成本价为2493元人民币，而在“中国智造”模式下成本可控制在100元人民币以内（图片来源：OMOTE 3D）

由此可见，即使在由“批量生产”转向“批量定制”的时代，以 3D 打印为代表的“第三次工业革命”仍有很大的希望在中国落地生根，形成“中国智造”的新模式，而不是制造业回流到欧美。3D 打印和智能数字化技术将深刻改变传统行业的产业模式，将为我国制造业的转型发展带来前所未有的机遇。

第2章

3D打印机的原理与种类

《礼记·大学》有云：“致知在格物，物格而后知至”。所谓“格物致知”，说得直白点，就是探究事物的实际原理，从而获得对它的真实认知。同样，对于 3D 打印而言，我们首先最重要的就是要明白它的工作原理。当前，3D 打印正处于蓬勃发展期，面对诱人的巨大市场前景，各个厂商纷纷提出各种新的技术工艺，让人有一种眼花缭乱的感觉。尤其对于初学者，往往觉得一团雾水。其实，只要我们掌握了 3D 打印的基本原理，那么“万变不离其宗”，冷冰冰的术语就不会再有距离感。

本章中，对每一项 3D 打印技术都通过图文并茂的方式展示出来。更重要的是，指出了每项技术的区别和联系，以及各自的优缺点，希望能带给读者一种“一览众山小”的轻松感觉。

2.1 3D打印时间简史——源自1860

如果寻根问底的话，3D 打印的发展最早可以追溯到 19 世纪！快速成型技术就是从这一时期开始萌芽的。但直到 20 世纪 80 年代后期，3D 打印技术才真正开始发展成熟并被广泛商业应用。

1860 年，法国人 François Willème 申请到了多照相机实体雕塑（Photosculpture）的专利。

1986 年，查尔斯·W·哈尔（Charles W. Hull，如图 2-1 所示）成立了世界上第一家生产 3D 打印设备的公司：3D Systems 公司。他研发了现在通用的 STL 文件格式。

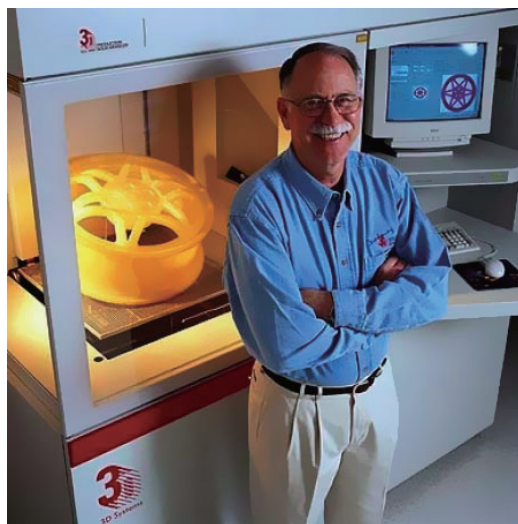


图 2-1 Charles W. Hull 成立了世界上第一家生产 3D 打印设备的公司（图片来源：3D Systems）

1988年，3D Systems公司在成立两年后，推出了世界上第一台基于SL（立体光刻）技术的3D工业级打印机SLA-250。同年，Scott Crump发明了另一种更廉价的3D打印技术：熔融沉积成型（FDM）技术，并于1989年成立了Stratasys公司。

1989年，美国得克萨斯大学奥斯汀分校的C. R. Dechard发明了选择性激光烧结工艺（SLS）。SLS使用的材料最广泛，理论上讲几乎所有的粉末材料都可以打印，如陶瓷、蜡、尼龙，甚至是金属。

1991年，Helisys推出第一台叠层法快速成型（LOM）系统。

1992年，Stratasys公司在成立3年后，推出了第一台基于FDM技术的3D工业级打印机。

1992年，DTM公司推出首台选择性激光烧结（SLS）打印机。

1993年，美国麻省理工学院MIT的Emanuel Sachs教授发明了三维打印技术（Three-Dimension Printing, 3DP），是类似于已在二维打印机中运用的喷墨打印技术。1995年，Z Corporation获得MIT的许可，并开始开发基于3DP技术的打印机。



注意：MIT发明的三维打印技术（Three-Dimension Printing, 3DP）只是“3D打印”众多成型技术中的一种而已。我们通常所说的“3D打印”并非特指MIT的这项3DP技术。

1996年，3D Systems、Stratasys、Z Corporation（以下简称ZCorp）各自推出了新一代的快速成型设备，此后快速成型便有了更加通俗的称呼——“3D打印”。

1998年，Optomec成功开发LENS激光烧结技术。

2000年，Objet更新SLA技术，使用紫外线光感和液滴喷射综合技术，大幅提高制造精度。

2001年，Solido开发出第一代桌面级3D打印机。

2003年，EOS开发DMLS激光烧结技术。

2005年，ZCorp公司推出世界上第一台高精度彩色3D打印机Spectrum Z510，让3D打印从此变得绚丽多彩。

2007年，3D打印服务创业公司Shapeways正式成立，Shapeways公司提供给用户一个个性化产品定制的网络平台。

2008年，第一款开源的桌面级3D打印机RepRap发布，其目的是开发一种能自我复制的3D打印机。RepRap是英国巴恩大学高级讲师Adrian Bowyer于2005年发起的开源3D打印机项目，如图2-2所示。该项目的目标是使工业生产变得大众化，全球各地的每个人都能以低成本打印RepRap的组装件，然后用打印机制造出日常用品。桌面级的开源3D打印机为轰轰烈烈的3D打印普及化浪潮揭开了序幕。



提示：值得一提的是，RepRap打印机创始人Adrian Bowyer之前的研究领域是3D数字化几何建模。

2008年，Objet Geometries公司推出其革命性的Connex500™快速成型系统，它是有史以来第一台能够同时使用几种不同的打印原料的3D打印机。

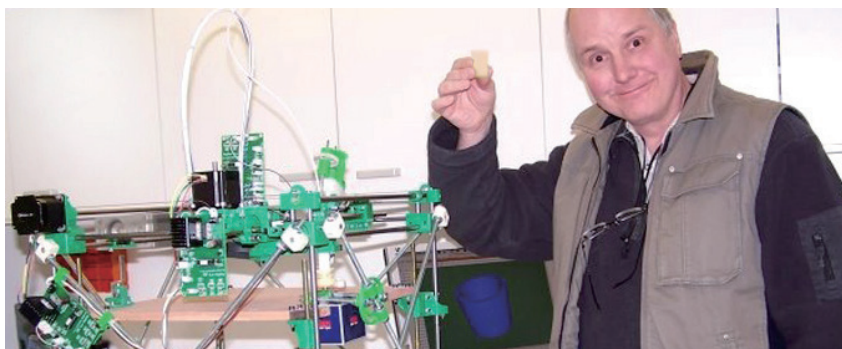


图 2-2 Adrian Bowyer 与 RepRap (图片来源 : RepRap)

2009 年, Bre Pettis 带领团队创立了著名的桌面级 3D 打印机公司——MakerBot, MakerBot 打印机源自于 RepRap 开源项目。MakerBot 出售 DIY 套件, 购买者可自行组装 3D 打印机。国内的创客开始了仿造工作, 个人 3D 打印机产品市场由此蓬勃兴起。

2010 年 12 月, Organovo 公司, 一个注重生物打印技术的再生医学研究公司, 公开第一个利用生物打印技术打印完整血管的数据资源。

2011 年, 英国南安普敦大学的工程师们设计和试驾了全球首架 3D 打印的飞机。这架无人飞机的建造用时 7 天, 费用为 5 000 英镑。3D 打印技术使得飞机能够采用椭圆形机翼, 有助于提高空气动力效率; 若采用普通技术制造此类机翼, 通常成本较高。

2011 年, Kor Ecologic 推出全球第一辆 3D 打印的汽车 Urbee。它是史上第一台用巨型 3D 打印机打印出整个身躯的汽车。所有外部组件也由 3D 打印制作完成。

2011 年 7 月, 英国研究人员开发出世界上第一台 3D 巧克力打印机。

2011 年, i.materialise 成为全球首家提供 14K 黄金和标准纯银材料打印的 3D 打印服务商。这在无形中为珠宝首饰设计师们提供了一个低成本的全新生产方式。

2012 年, 荷兰医生和工程师们使用 LayerWise 制造的 3D 打印机, 打印出一个定制的下颌假体。然后移植到一位 83 岁的老太太身上。这位老太太患有慢性骨感染。目前, 该技术被用于促进新的骨组织生长。

2012 年, 英国著名经济学杂志《经济学人》封面文章 (如图 2-3 所示), 声称 3D 打印将引发全球第三次工业革命。

2012 年 3 月, 维也纳大学的研究人员宣布利用双光子光刻 (Two-Photon Lithography) 突破了 3D 打印的最小极限, 展示了一辆不到 0.3mm 的赛车模型。



图 2-3 《经济学人》的封面文章
《The third industrial revolution》

2012年3月，美国总统奥巴马提出投资10亿美元在全美建立15家制造业创新研究所。

2012年7月，比利时的International University College Leuven的一个研究组测试了一辆几乎完全由3D打印的小型赛车。车速达到了140km/h。

2012年9月，3D打印的两个领先企业Stratasys和以色列的Objet宣布进行合并，合并后的公司名仍为Stratasys，进一步确立了Stratasys在高速发展的3D打印及数字制造业中的领导地位。

2012年10月，来自MIT的团队成立Formlabs公司，并发布了世界上第一台廉价且高精度的SLA个人3D打印机Form 1。国内的创客也由此开始研发基于SLA技术的个人3D打印机。

同期，中国3D打印技术产业联盟正式宣告成立。国内各类媒体开始铺天盖地报道3D打印的新闻。

2012年11月，中国宣布是世界上唯一掌握大型结构关键件激光成型技术的国家。

2012年11月，苏格兰科学家利用人体细胞首次用3D打印机打印出人造肝脏组织。

2013年5月，美国分布式防御组织发布全世界第一款完全通过3D打印制造出的塑料手枪（除了撞针采用金属），并成功试射。同年11月，美国Solid Concepts公司制造了全球第一款3D全金属手枪，采用33个17-4不锈钢部件和625个铬镍铁合金部件制成，并成功发射了50发子弹。

2013年，美国的两位创客（父子俩）开发出家用金属3D打印机，基于液体金属喷射打印（LMJP）工艺，价格将低于10 000美元。同年，美国的另外一个创客团队开发了一款名为Mini Metal Maker（小型金属制作者）的桌面级金属3D打印机，主要打印一些小型的金属制品，比如珠宝、金属链、装饰品、小型金属零件等，售价仅为1 000美元。

2013年8月，美国国家航空航天局（NASA）测试3D打印的火箭部件，其可承受2万磅推力，并可耐6 000华氏度的高温。

2013年，麦肯锡公司将3D打印列为12项颠覆性技术之一，并预测到2025年，3D打印对全球经济的价值贡献将为2~6千亿美元。

2014年7月，美国南达科塔州一家名为Flexible Robotic Environments（FRE）的公司公布了最新开发的全功能制造设备VDK6000，兼具金属3D打印（增材制造）、车床（减材制造，包括：铣削、激光扫描、超声波检具、等离子焊接、研磨/抛光/钻孔）及3D扫描功能。

2014年8月，国外一名年仅22岁的创客Yvo de Haas推出了3DP工艺的桌面级3D打印机Plan B，技术细节完全开源，自己组装费用仅需1 000欧元。

2014年10月，国外3名创客成立的Sintratec公司，推出了一款SLS工艺的3D打印机，售价仅为3 999欧元。

2.2 3D打印机的工作原理和家族

3D 打印，又称快速成型（RP，Rapid Prototyping）、增材制造（AM，Additive Manufacturing），是一种以 3D 数字模型文件为基础，运用粉末状金属或塑料等可黏结材料，通过逐层打印的方式来构造物体的技术。在将 3D 数字化模型输出到 3D 打印机之前，需要对 3D 模型进行分层，切成数百上千个薄层，这相当于高等数学里的微分操作。然后将描述这些薄层的数字化文件输出到打印机，3D 打印机逐层打印出来，这又相当于高等数学里的积分操作，直到将整个形状叠加成型。

2.2.1 3D 打印机的工作原理与流程

3D 打印采用分层加工、叠加成型，即通过逐层增加材料来生成 3D 实体，与传统的去除材料加工技术（如用机床切削）完全不同。之所以称之为“打印机”，是因为分层加工的过程与喷墨打印十分相似，组成上也都是由控制组件、机械组件、打印头、耗材和介质等构成的。

说得简单一点，3D 打印是断层扫描的逆过程，断层扫描是把某个东西“切”成无数叠加的片，3D 打印就是一片一片地打印，然后叠加到一起，成为一个立体物体，如图 2-4 所示。在 3D 打印时，软件通过计算机辅助设计技术（CAD）完成一系列数字“切片”（Slice），并将这些切片的信息传送到 3D 打印机上，然后将连续的薄型层面堆叠起来，直到一个固态物体成型。3D 打印机与传统打印机最大的区别在于它使用的“墨水”是实实在在的原材料。

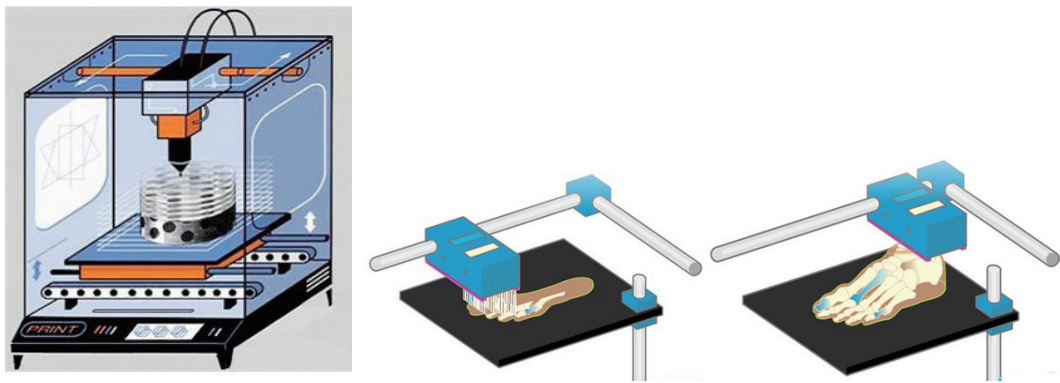


图 2-4 3D 打印机的“分层加工、叠加成型”工作原理图（图片来源：blogspot）

传统的“去材加工”机床是在做“减法”（**减材成型**），也即通过切、削、车、铣、磨等工艺将一块物料上不需要的地方去掉，但这就存在着“伸不进、够不着”的问题，因此不能加工任意复杂的中空形状，而且去掉的物料也被浪费掉了。作为对比，3D 打印这种一层一层堆积起来做“加法”的工艺（**增材成型**）具有如下优点：不需要刀具、模具，所需工装、夹具大幅度减少；生产周期大幅度缩短；可制造出传统工艺方法难以加工，甚至无法加工的结构；材料利用率大幅度提高。因此，3D 打印特别适合于复杂结构的快速制造、个性化定制、高附加值的产品制造。同时，由于可以生成任意复杂的产品形状，因此在零部件的设计上可以采用最优的结构设计，而无须考虑加工问题，解决了复杂精细零部件的设计和制造难题。



提示：因为 3D 打印有 X 、 Y 、 Z 这 3 个空间维度，所以满足所谓的“三次方增加规律”。比如物品的每个维度（长、宽、高）都增大到 2 倍，则体积将增大到 $2^3=8$ 倍！也即从时间、材料到成本的需求数量都是呈指数增长的，有时甚至能达到三次方增长。所以说，如果我们需要 2 倍大的东西，那么就得分花 8 倍的时间、花 8 倍的材料、花 8 倍的钱来打印。

3D 打印的主流技术包括 SLA、FDM、SLS、3DP、LOM 等。比如，FDM 是把塑料熔化成半融状态拉成丝，用线来构建面，一层一层堆起来；而光固化 SLA 是把本来液态的光敏树脂，用紫外激光照射，照到哪儿，哪儿就从液态变成了固态。SLS 和 SLA 理论上是一样的，不同的是 SLS 用激光去烧结粉末，如尼龙粉、金属粉等。

下面，我们就对各种主流 3D 打印技术做一个通俗易懂的介绍。

2.2.2 FDM：熔融沉积成型（FFF：熔丝制造）

FDM（Fused Deposition Modeling），熔融沉积成型。因为 FDM 已被 Stratasys 注册商标，所以其他厂商将其改称为熔丝制造（Fused Filament Fabrication，FFF）、塑料喷印（PJP）、熔丝建模（FFM）等。该工艺属于“丝材挤出热熔成型”这一大类。

FDM 技术是 20 世纪 80 年代时 Scott Crump 发明的。在获得该项技术的专利后，他于 1989 年建立了 Stratasys 公司。FDM 的技术原理是，如图 2-5 所示，将丝状（直径约 2mm）的热塑性材料通过喷头加热熔化，喷头底部带有微细喷嘴（直径一般为 0.2 ~ 0.6 mm），材料以一定的压力挤喷出来，同时喷头沿水平方向移动，挤出的材料与前一个层面熔结在一起。一个层面沉积完成后，工作台垂直下降一个层的厚度，再继续熔融沉积，直至完成整个实体造型。FDM 工艺使用两种材料：一种是制作实体部分的成型材料；另一种是支撑材料，以防空腔或悬臂部分坍塌。

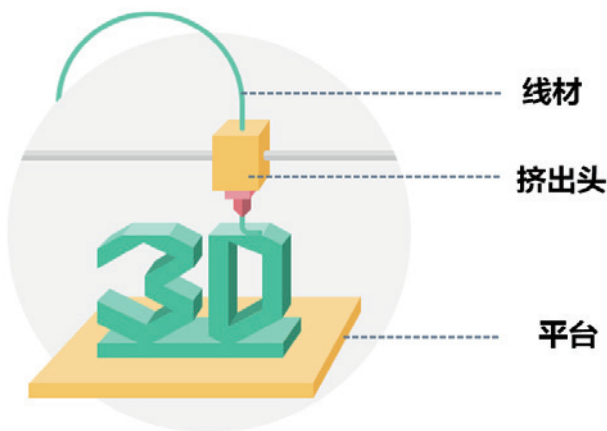


图 2-5 FDM 的技术原理（图片来源：thre3d.com）

形象地说，FDM 的原理就像蚕吐丝或挤牙膏那么简单，且无须激光系统，因而价格低廉。现在市场上的桌面级 3D 打印机（如 RepRap、Ultimaker、MakerBot）大多数采用这种工艺，最便宜的不到 1 万元即可买到。FDM 使用的丝状耗材以及打印案例如图 2-6 所示。

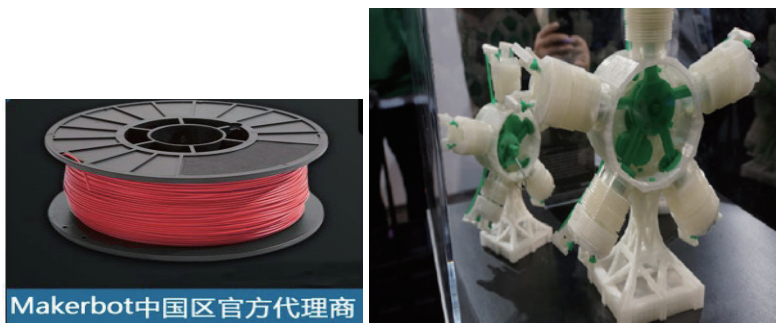


图 2-6 FDM 使用的丝状耗材以及打印案例（图片来源：MakerBot）

FDM 技术的优点如下。

- ✓ 操作环境干净、安全，可在办公室环境下进行，没有产生毒气和化学污染的危险。
- ✓ 无须激光器等贵重元器件，工艺简单、干净、不产生垃圾。
- ✓ 原材料以卷轴丝的形式提供，易于搬运和快速更换。
- ✓ 材料利用率高，且可选用多种材料，如可染色的 ABS 和医用 ABS、PLA、PC、PPSF 等。
- ✓ 由于甲基丙烯酸 ABS（MABS）材料具有较好的化学稳定性，可采用伽马射线消毒，特别适用于医用。

FDM 技术的缺点如下。

- ✗ 成型后表面粗糙，需配合后续抛光处理，目前不适合高精度的应用。做小件或精细件时精度不如 SLA，最高精度只能为 0.1mm。
- ✗ 尺寸不能很大，因为材料本身原因限制，尺寸大了很容易变形。
- ✗ 速度较慢，因为它的喷头是机械的。
- ✗ 此外它还需要浪费材料来做支撑。



提示：在成型过程中，丝材经小孔挤出时，喷嘴喷出的熔丝会在出口区域形成“膨化现象”，也即填充的实际轮廓线要超出理论轮廓线。具体地，挤出丝的实际**线宽** w 同时受到喷嘴直径 d 、分层厚度 δ 、挤出速度 v_e 、扫描速度 v_s 等多方面因素的影响，根据流入/流出体积守恒的原理（假设不考虑材料收缩），可知：

$$w = (v_e \pi d^2) / (4v_s \delta)$$

由上式可见，若扫描速度 v_s 不变，**随着挤出速度 v_e 的增大，线宽 w 逐渐增大**。特别是当挤出速度超出一定范围后，挤出丝就会黏附在喷嘴的外表面，从而影响打印。因此，扫描速度须与挤出速度相匹配。根据经验，线宽 w 一般设置为喷嘴直径的 1.3~1.6 倍。此外，3D 打印与传统加工方法相比，具有微观非均匀性及层性、各向异性、性能蠕变等特性，因此理论研究涉及材料力学、弹性力学等领域。比如，我们要研究 3D 打印零件的变形，则需考虑层内**应力—应变分布**、层间应力—应变分布等方面。以层内应力分析为例，复合材料理论认为：固化物可视为弹性体，因此当应力低于弹性极限时，其应力—应变关系满足**广义胡克定律**，也即应变分量是应力分量的线性函数，

$$\begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{yz} \\ \gamma_{zx} \\ \gamma_{xy} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} & S_{16} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} & S_{26} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} & S_{36} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} & S_{56} \\ S_{61} & S_{62} & S_{63} & S_{64} & S_{65} & S_{66} \end{pmatrix} \begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{yz} \\ \tau_{zx} \\ \tau_{xy} \end{pmatrix}$$

上式简写为：

$$\boldsymbol{\Gamma} = \boldsymbol{S}\boldsymbol{\Phi}$$

其中 $\boldsymbol{\Gamma}$ 为**应变矩阵**； \boldsymbol{S} 为**柔度矩阵**，其逆矩阵称为刚度矩阵， $S_{ij}(i, j=1, 2, \dots, 6)$ 代表弹性体的弹性系数，也即单位力作用下产生的变形量； $\boldsymbol{\Phi}$ 为**应力矩阵**。

2.2.3 3DP：三维打印黏结成型（喷墨沉积）

3DP（Three Dimensional Printing and Gluing），三维打印黏结成型、喷墨沉积，也被称为黏合喷射（Binder Jetting）、喷墨粉末打印（Inkjet Powder Printing）。该工艺属于“液体喷印成型”这一大类。

工艺类似于传统的 2D 喷墨打印机，是最为贴合“3D 打印”概念的成型技术之一。最早由美国麻省理工学院（MIT）于 1993 年开发。该技术利用喷头喷黏结剂，选择性地黏结粉未来成型。如图 2-7 所示，首先铺粉机构在加工平台上精确地铺上一薄层粉末材料，然后喷墨打印头根据这一层的截面形状在粉末上喷出一层特殊的胶水，喷到胶水的薄层粉末发生固化。然后在这一层上再铺上一层一定厚度的粉末，打印头按下一截面的形状喷胶水。如此层层叠加，从下到上，直到把一个零件的所有层打印完毕。然后把未固化的粉末清理掉，得到一个三维实物原型，成型精度可达 0.09mm。因为石膏成型品十分易碎，因此后期还可采用“浸渍”处理，比如采用盐水或加固胶水（Z-Bond、Z-Max 等），使之变得坚硬。

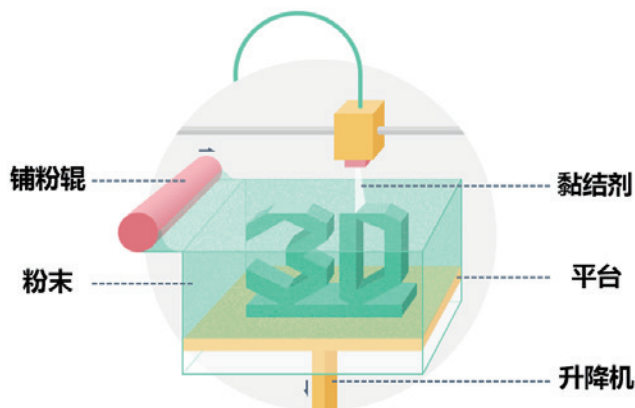


图 2-7 3DP 的技术原理（图片来源：thre3d.com）

与 2D 平面打印机在打印头下送纸不同，3D 打印机是在一层粉末的上方移动打印头，并打

印横截面数据。彩色 3D 打印机打印成型的样品模型与实际产品具有同样丰富的色彩。ZCorp 公司（现已被 3D Systems 公司收购）使用 3DP 打印技术开发了 Zprinter 产品系列。3DP 的打印案例如图 2-8 所示。



图 2-8 3DP 的打印案例（图片来源：3D Systems）

3DP 技术的优点如下。

- ✓ 无须激光器等高成本元器件。成型速度非常快（相比于 FDM 和 SLA），耗材很便宜，一般的石膏粉都可以。
- ✓ 成型过程不需要支撑，多余粉末的去除比较方便，特别适合于做内腔复杂的原型。
- ✓ 此技术最大优点是能直接打印彩色，无须后期上色。目前市面上打印彩色人像基本采用此技术。

3DP 技术的缺点如下。

- ✗ 石膏强度较低，只能做概念型模型，而不能做功能性试验。
- ✗ 因为是粉末黏结在一起，所以表面手感稍有些粗糙。



提示：除了 3DP（黏结剂 + 石膏粉），还有其他的黏结剂喷射技术，比如 3D 砂型铸造、黏结剂喷射金属打印、黏结剂喷射陶瓷打印、黏结剂喷射玻璃打印等，采用的都是类似的原理，只不过选择的原材料不同。

2.2.4 SLS：选择性激光烧结

SLS（Selective Laser Sintering），选择性激光烧结、选区激光烧结。该工艺属于“粉末 / 丝状材料高能束烧结或熔化成型”这一大类。

该工艺由美国得克萨斯大学奥斯汀分校的 C. R. Dechard 于 1989 年研制成功。SLS 与 3DP 相似，也是采用粉末材料，但一般都为金属粉末、陶瓷粉末等。此外，不像 3DP 通过喷头喷黏结剂来黏结，而是通过烧结来黏结。具体地，如图 2-9 所示，SLS 利用粉末材料在激光照射下烧结的原理，由计算机控制，层层堆结成型。首先铺一层粉末材料，并刮平。将材料预热到接近熔化点，再使用高强度的 CO₂ 激光器有选择地在该层截面上扫描，使粉末温度升至熔化点，然

后烧结形成黏结，接着不断重复铺粉、烧结的过程，直至完成整个模型成型。

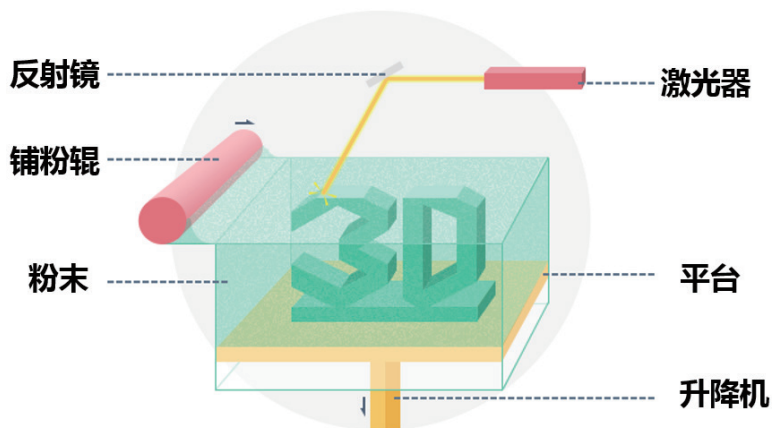


图 2-9 SLS 的技术原理（图片来源：thre3d.com）

SLS 在市场上采用得比较多，因为它和工业结合得很紧密，而且使用的材料最广泛，理论上讲几乎所有的粉末材料都可以打印。像铸造行业对精度要求没那么高，SLS 打印出来的精度足够了，与精密铸造工艺相当。SLS 可以直接打印一些小的金属件，如首饰、小的金属模具等。SLS 的打印案例如图 2-10 所示。



图 2-10 SLS 的打印案例（图片来源：ArtCorp）

SLS 技术的优点如下。

- ✓ 成型材料广泛，包括高分子、金属、陶瓷、砂等多种粉末材料。
- ✓ 零件的构建时间较短，可达到 1 inch/h 速度。
- ✓ 所有没用过的粉末都能在下次打印中循环利用。所有未烧结过的粉末都保持原状并成为实物的支撑性结构，因此这种方法不需要任何其他支撑材料。相比之下，FDM、SLA 等工艺则需要支撑结构。
- ✓ 此技术最主要的优势在于金属成品的制作，其制成的产品可具有与金属零件相近的机械性能，故可用于直接制造金属模具以及进行小批量零件生产。

SLS 技术的缺点如下。

- ✗ 粉末烧结的表面粗糙（精度为 0.1 ~ 0.2mm），需要后期处理。在后期处理中难以保证

制件尺寸精度，后期处理工艺复杂，样件变形大，无法装配。

- ✗ 无法直接成型高性能的金属和陶瓷零件，成型大尺寸零件时容易发生翘曲变形。
- ✗ 在加工前，要花近 2 小时的时间将粉末加热到熔点以下，当零件构建之后，还要花 5 ~ 10 小时冷却，然后才能将零件从粉末缸中取出。
- ✗ 由于使用了大功率激光器，除了本身的设备成本，还需要很多辅助保护工艺，整体技术难度较大，制造和维护成本非常高，普通用户无法承受，所以目前应用范围主要集中在高端制造领域，而尚未有桌面级 SLS 打印机开发的新闻。
- ✗ 需要对加工室不断充氮气以确保烧结过程的安全性，加工的成本高。该工艺产生有毒气体，污染环境。

还有一种跟 SLS 原理相似的工艺名为 SHS (Selective Heat Sintering : **选择性热烧结**)，如图 2-11 所示。不同之处在于 SHS 采用的是热打印头而非高强度的激光，耗材为热塑性粉末而非金属粉末，因此是一种相对廉价的方案，可被用于桌面级打印。

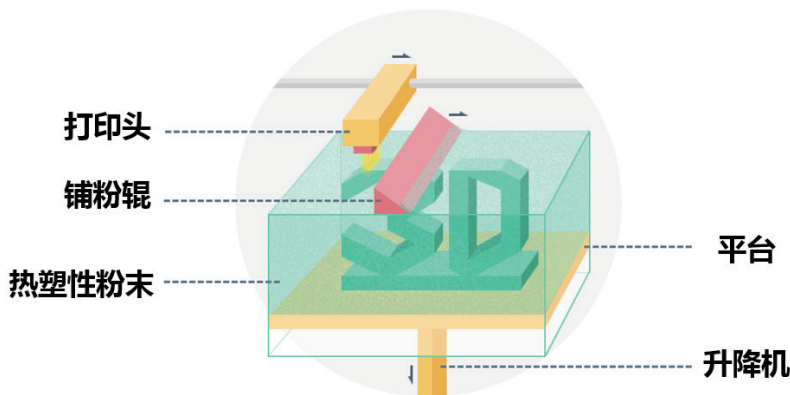


图 2-11 SHS 的技术原理 (图片来源 : thre3d.com)

除了 SLS，金属打印还有 SLM、DMLS、LENS、EBM、EBDM，详细介绍请移步本章 2.4 节。

2.2.5 SLA : 光固化立体成型 (立体光刻)

SLA (Stereo Lithography Appearance)，光固化立体成型、立体光刻、立体平板印刷，有时也简称 SL。该工艺属于“液态树脂光固化成型”这一大类。



提示：SLA 用的激光与 SLS 用的激光不同。SLA 用的是紫外激光，而 SLS 用的是红外激光。SLA 的耗材一般为液态的光敏树脂，而 SLS 的耗材一般为塑料、蜡、陶瓷、金属粉末。

世界上第一台 3D 打印机采用的就是 SLA 工艺！这项技术由 Charles W. Hull 发明，他由此于 1986 年创办了 3D Systems 公司。技术原理：如图 2-12 所示，树脂液槽中盛满透明、有黏性的液态光敏树脂，紫外激光束经快速转动着的反射镜（即**振镜**）对树脂进行照射，使之快速固化。具体地，在成型过程开始时，可升降的工作台处于液面下一个截面层厚的高度。之后，聚焦的激光束在计算机的控制下，按照截面轮廓的要求，沿液面进行扫描，使被扫描区域的树脂固化，

从而得到该截面轮廓的塑料薄片。然后，工作台下降一层薄片的高度，再固化另一个层面。这样层层叠加构成一个三维实体。

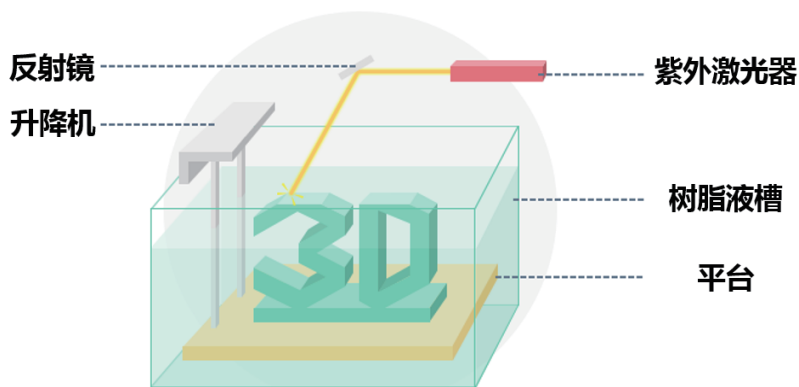


图 2-12 SLA 的技术原理（图片来源：thre3d.com）

SLA 的材料是液态的，不存在颗粒的东西，因此可以做得很精细。不过它的材料要比 SLS 贵很多，所以它目前主要用于打印薄壁的、精度要求较高的零件。适合于制作中小型工件，能直接得到塑料产品。它还能代替蜡模制作浇铸模具，以及作为金属喷涂模、环氧树脂模和其他软模的母模。SLA 的打印案例如图 2-13 所示。



图 2-13 SLA 的打印案例（图片来源：中瑞科技）

SLA 技术的优点如下。

- ✓ 光固化成型法是最早出现的快速成型制造工艺，成熟度最高，经过时间的检验。
- ✓ 成型速度较快，系统工作相对稳定。
- ✓ 可以打印的尺寸也比较可观，在国外有可以做到 2m 的大件，关于后期处理特别是上色都比较容易。
- ✓ 尺寸精度高，可以做到微米级别，比如 0.025mm。
- ✓ 表面质量较好，比较适合做小件及较精细件。

SLA 技术的缺点如下。

- ✗ SLA 设备造价高昂，使用和维护成本过高。SLA 系统是要对液体进行操作的精密设备，对工作环境要求苛刻。
- ✗ 成型件多为树脂类，材料价格贵，强度、刚度、耐热性有限，不利于长时间保存。
- ✗ 这种成型产品对贮藏环境有很高的要求，温度过高会熔化，工作温度不能超过 100℃。光敏树脂固化后较脆，易断裂，可加工性不好。成型件易吸湿膨胀，抗腐蚀能力不强。
- ✗ 光敏树脂对环境有污染，会使人体皮肤过敏。
- ✗ 需要设计工件的支撑结构，以便确保在成型过程中制作的每一个结构部位都能可靠定位，支撑结构需在未完全固化时手工去除，容易破坏成型件。

2.2.6 PolyJet : 多头喷射技术 (Material Jetting : 材料喷射)

PolyJet 技术是由以色列 Objet 公司 (现已并入 Stratasys 公司) 发明并申请专利的，速度比 SLA 更快。该工艺属于“液体喷印成型”和“液态树脂光固化成型”这两大类的结合体。

打印过程像喷墨打印机一样一层一层地喷树脂，如图 2-14 所示，同时用紫外线灯快速固化，树脂分为支撑材料和模型材料，产品做成后可轻易地冲洗掉支撑材料产品。样品精度最高可达到 16mm。Objet 的 3D 打印系统系列可为办公室环境的设计师和工程师们提供高分辨率的快速原型制作。缺点是树脂材料的强度较低。

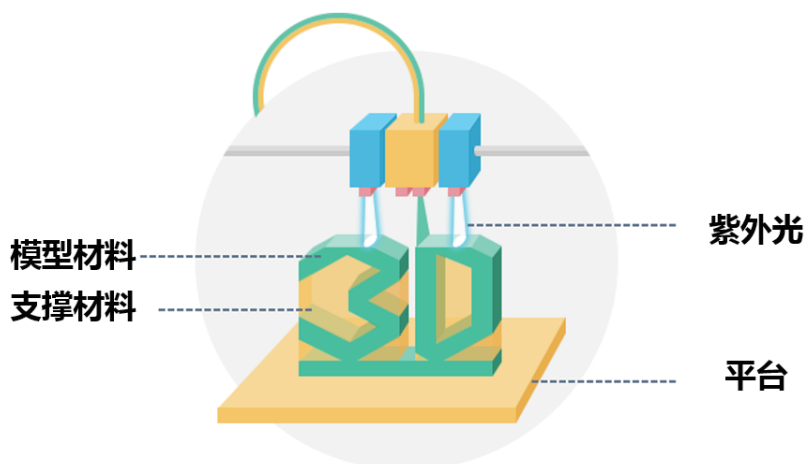


图 2-14 PolyJet 的技术原理 (图片来源 : thre3d.com)

此外，如果搭配 FullCure720 耗材，可实现透明的琥珀色效果。Objet 最新的 PolyJet Matrix 技术，还可以支持多种型号材料 (多种颜色) 同时喷射。PolyJet 工艺的打印案例如图 2-15 所示。



图 2-15 PolyJet 工艺的打印案例（图片来源：Stratasys）

2.2.7 DLP：数字光处理

DLP（Digital Light Processing, 数字光处理技术）也属于“液态树脂光固化成型”这一大类，数字光处理技术和 SLA 光固化成型技术比较相似，不过它是使用高分辨率的数字光处理器（DLP）投影仪来固化液态光聚合物的，逐层地进行光固化，如图 2-16 所示。由于每次成型一个面，因此在理论上速度也比同类的 SLA 快很多。

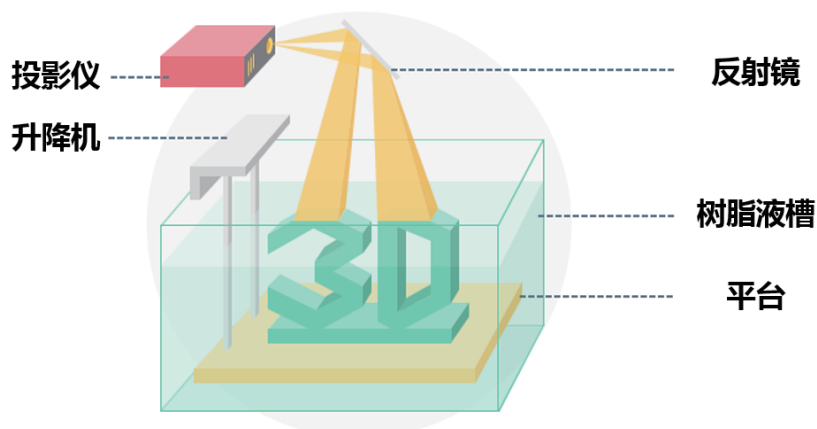


图 2-16 DLP 的技术原理（图片来源：thre3d.com）

该技术成型精度高，在材料属性、细节和表面光洁度方面可匹敌注塑成型的耐用塑料部件。DLP 工艺的打印案例如图 2-17 所示。

DLP 利用投射原理成型，无论工件尺寸大小都不会改变成型速度。此外，DLP 不需要激光头去固化成型，取而代之是使用成本极为便宜的灯泡照射。整个系统并没有喷射部分，所以并没有传统成型系统喷头堵塞的问题出现，大大降低了维护成本。DLP 技术最早由德州仪器开发，目前很多产品也是基于德州仪器提供的芯片组。



图 2-17 EnvisionTEC 公司 Perfactory 的打印案例，采用 RC31 材料（图片来源：envisionTEC）

ZCorp 公司使用 DLP 技术开发了 ZBuilder 产品系列，使得工程师能够在产品大规模生产前验证设计的形状、匹配和功能，从而避免成本高昂的生产模具修改和缩短上市时间。

有一个好消息是，国外一名叫 Tristram Budel 的创客发布了一款开源的高分辨率 DLP 3D 桌面打印机（如图 2-18 所示），所有技术细节都免费共享。此外，美国创客 Michael Joyce 发起的 B9Creator 开源项目目前在市场上获得了较大的成功。

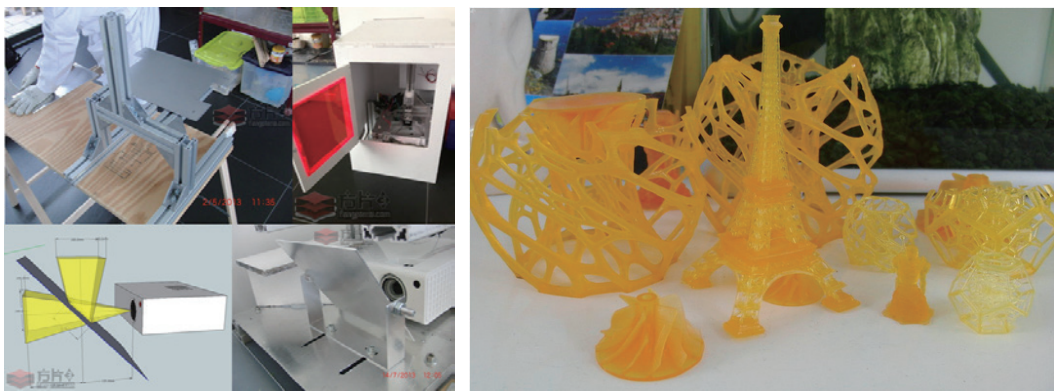


图 2-18 开源的高分辨率 DLP 3D 桌面打印机（图片来源：Tristram Budel）



提示：除了 SLA 和 DLP，还有一种也是选择性固化液体的 3D 打印技术：**双光子光刻 (2PP)**，效果图片见第 1 章 1.1 节。这是一种纳米级的 3D 打印技术，未来可能会成为主流的 3D 打印形式。如果说传统的光固化技术可以达到微米级别，比如 0.025mm，那么 2PP 则可达**纳米级别**，所有轴的分辨率都达到了 0.000 1mm，也即精准了 250 倍，所打印出的物品比细菌还要小。同时，2PP 打印速度极快，每秒可打印几米的物体。

2.2.8 LOM：分层实体制造

LOM (Laminated Object Manufacturing)，分层实体制造。该工艺属于“片 / 板 / 块材黏接

或焊接成型”这一大类。

LOM 是一种薄片材料叠加工艺，出现于 1986 年，由 Helisys 公司提出。如图 2-19 所示，利用激光或刀具切割薄片纸、塑料薄膜、金属薄板或陶瓷薄片等片材，非零件区域切割成若干小方格，便于后续去除。然后通过热压或其他形式层层黏结，叠加获得三维实体零件。可以看出，LOM 工艺还有传统切削工艺的影子，只不过它已不是对大块原材料进行整体切削，而是先将原材料分割为多层，然后对每层的内外轮廓进行切削加工成型，并将各层黏结在一起。

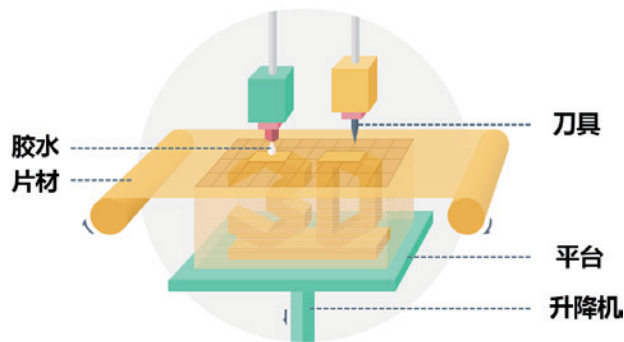


图 2-19 LOM 的技术原理（图片来源：thre3d.com）

LOM 适合制作大中型原型件，翘曲变形较小，尺寸精度较高，成型时间较短。使用小功率 CO₂ 激光器价格低、使用寿命长，制成件有良好的机械性能，适合于产品设计的概念建模和功能性测试零件。且由于制成的零件具有木质属性（激光切割能使纸张边缘轻微变成棕色），特别适合于直接制作砂型铸造模。LOM 原理图和打印案例如图 2-20 所示。



图 2-20 上图：LOM 原理图（左）和打印案例（中、右）。下图：去除 LOM 的支撑结构，取出成型件（图片来源：rpworld.net）

LOM 技术的优点如下。

- ✓ 成本低；因为没有涉及化学反应，所以零件可做得很大。

- ✓ 仅切割内外轮廓，内部无须加工，所以这是一个高速的快速成型工艺。常用于加工内部结构简单的大型零件及实体件。
- ✓ 不存在收缩和翘曲变形，无须设计和构建支撑结构。

LOM 技术的缺点如下。

- ✗ 不能制造中空结构件。难以构建精细形状的零件，即仅限于结构简单的零件。
- ✗ 比较浪费材料。可实际应用的原材料种类较少，如纸、塑料、陶土以及合成材料，但目前常用的只是纸。
- ✗ Z轴精度比 SLA 低，精度可达 0.1mm；且纸制零件很容易吸潮，必须立即进行后处理、上漆。
- ✗ 需要专门实验室环境，维护费用高昂。当加工室的温度过高时常有火灾发生。因此，工作过程中需要专职人员职守。

2.3 塑料还是石膏？3D打印机的各种耗材

提到 3D 打印，很多人都会问一个问题：3D 打印技术都能使用哪些材料呢？其实想回答这个问题，我们只需要看看欧美最著名的三大主要 3D 打印服务公司 Shapeways、i.materialise 和 Ponoko 目前都正在为客户们提供哪些材料，就可以大致了解当今世界都有哪些成熟的 3D 打印材料了。

在将各种耗材“一网打尽”之前，我们先挑出几种应用最广泛的耗材专门详细介绍一下，比较它们之间的差别。特别是，目前 3D 照相馆的 3 种主要耗材为 ABS、PLA 和石膏，为此我们专门配上了各自的人像打印效果图。

ABS：丙烯腈-丁二烯-苯乙烯

ABS 塑料丝，五大合成树脂之一，是目前产量最大、应用最广泛的聚合物。它具有无毒、无味，价格便宜等特点。由于这种塑料丝要经过熔化后再冷却，根据热胀冷缩原理，所以在精度控制上并不是很高，打印的 3D 模型也比较粗糙。打印较大模型时，最好使用**热床**，否则产品容易起翘。ABS 在强度上高于 PLA。ABS 可以用丙酮进行后期打磨抛光。ABS 的打印案例如图 2-21 所示。

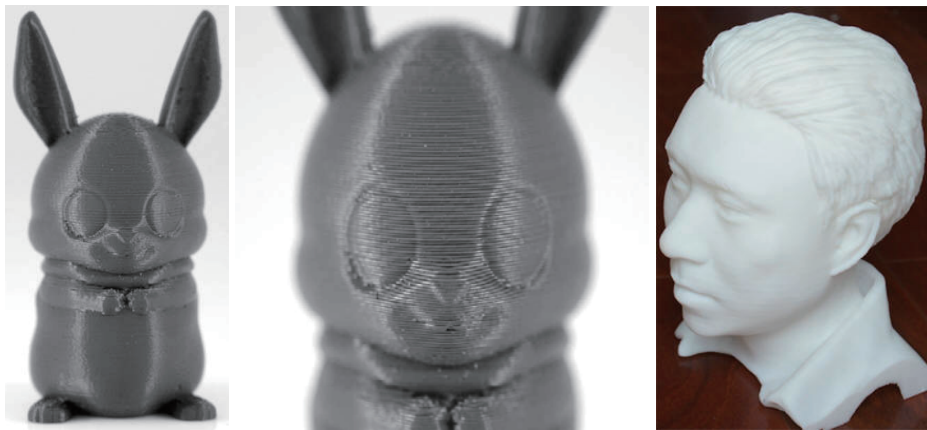


图 2-21 ABS 的打印案例（最右为笔者本人）

ABS 材料的特点如下。

- 综合性能较好，冲击强度较高，化学稳定性、电性能良好。
- 与 372 有机玻璃的熔结性良好，制成双色塑件，且可表面镀铬，喷漆处理。
- 有高抗冲、高耐热、阻燃、增强、透明等级别。
- 柔韧性好。
- 打印大尺寸模型时，模型容易变形翘曲。



注意：一般工厂都有防护措施，但我们使用家庭 3D 打印机的时候一般没有防护措施。美国和法国科研人员发表了一项科研成果，表明桌面 3D 打印机确实能够散发出许多的超细颗粒（UFPs, ultrafine particles），即那些小于 100nm 的微小颗粒。超细颗粒通常都很容易进入到人体的气管和肺部，并被吸收到血液循环系统，因而长期接触就有可能引发一些健康问题，比如肺病、中风和哮喘等。研究者发现，打印 ABS 的颗粒排放量为 PLA 的 10 倍。

PLA：聚乳酸，植物淀粉的衍生物

聚乳酸（PLA）是一种优良的聚合物，原因是多方面的。首先，它是生物环保的，因为它从玉米制成，是一种可再生资源。其次，生物可降解，这意味着将不需要一个垃圾填埋场。第三，它的颜色，打印的效果如 LED 一样地晶亮，非常清晰。最后，它具有极低的收缩率，这意味着它抗变形翘曲，即使是非常大的打印尺寸。PLA 的打印案例如图 2-22 所示。



图 2-22 PLA 的打印案例（笔者本人）

PLA 的很多性质与 ABS 类似，但比 ABS 更脆一点，无须使用加热平台，因为它冷却时很少发生卷翘，风扇可进一步提高打印质量。气味比 ABS 好闻。从表面上对比观察，ABS 呈亚光，而 PLA 很光亮。加热到 195℃，PLA 可以顺畅挤出，ABS 不可以。加热到 220℃，ABS 可以顺畅挤出，PLA 则会出现鼓起的气泡，甚至被碳化，碳化会堵住喷嘴。

打印 PLA 与打印 ABS 的区别如下。

- 打印 PLA 时有棉花糖气味，不像 ABS 那样有刺鼻的不良气味。
- 打印大型零件模型，PLA 即使在没有加热床的情况下边角也不会翘起。

- PLA 加工温度是 200℃，ABS 在 220℃以上。
- PLA 具有较低的熔体强度，打印模型更容易成型，表面光泽性优异，色彩艳丽。

在打印过程中，一般都需要对悬垂部分进行临时支撑。但这些支撑结构往往并不容易清除掉，用锉刀和砂纸也容易破坏光滑的表面。因此，除了 ABS 和 PLA 塑料，还有一种**水溶性的塑料 PVA**，可用在双喷头 3D 打印机上（比如一个喷头接 ABS 耗材以打印主体，另一个喷头接 PVA 耗材以打印支撑）。**PVA 可作为打印过程中的临时支撑材料来使用，打印完毕后泡在水基清洗剂中进行溶解，然后被很方便地清理掉。**

石膏粉末

使用粉末微粒作为打印介质（最常用的是石膏粉）的 3D 打印机，通过在粉末床上层层添加黏结剂的方式来成型。打印的模型较为精细，可再添加一层氰基丙烯酸酯密封胶来增加产品的耐用度并获得更鲜艳的色彩。非常适合做人像全彩打印，目前市面上很多 3D 照相馆用的就是石膏，但“不防水，不可放入洗碗机，不能回收再利用，对食品不安全，耐热度 60℃”。石膏粉末的打印案例如图 2-23 所示。



图 2-23 石膏粉末的打印案例（右边为笔者本人）

PC：聚碳酸酯

翘曲度低，强度高且比 ABS 更加柔韧一些。不过，挤出温度要求比 PLA 和 ABS 更高，有些个人 3D 打印机达不到。

尼龙

更柔韧，不需要加热平台或风扇就可获得细节度较高的打印表面光洁度。尼龙的打印案例如图 2-24 所示。



图 2-24 尼龙的打印案例（图片来源：magicfirm）

光敏树脂

光敏树脂是一种遇紫外线照射会立刻变硬的特殊材料。其特点是细节度和光滑度非常地高，非常适合对细节度要求高的雕塑和其他物品。光敏树脂的打印案例如图 2-25 所示。



图 2-25 光敏树脂的打印案例

纸张

Mcor 公司的 3D 打印机利用普通 A4 纸张打印全彩模型。

彩色和木质耗材

为了实现彩色打印，除了在打印机上想办法之外，还可在耗材上作文章。很简单的一个思路是，把耗材线染成彩色，这不就 OK 了？Taulman 的“618”高强度尼龙恰有这个优点，它很容易被染色，如图 2-26 上方所示是所打印的彩色效果。除了彩色耗材，木匠们可能更喜欢木质耗材。德国设计师 Kai Parthy 研制了一种名为 Laywoo-D3 的 3D 打印耗材，直径 3mm，其中 40% 源自回收木材，3D 打印完成后效果和胶合板类似，手感接近实木，并散发出木头的味道，如图 2-26 下方所示。目前国内的广州傲趣科技公司也推出一款名为 Pop Wood 的木质线材，价格便宜了许多，据报道说还不容易堵塞喷头。



图 2-26 彩色耗材（上）和木质耗材（下）（图片来源：Richrap、Kai Parthy）

好了，介绍完以上几种常见的耗材，下面我们将其他耗材进行系统的归类描述。它们的打印效果如图 2-27 所示。

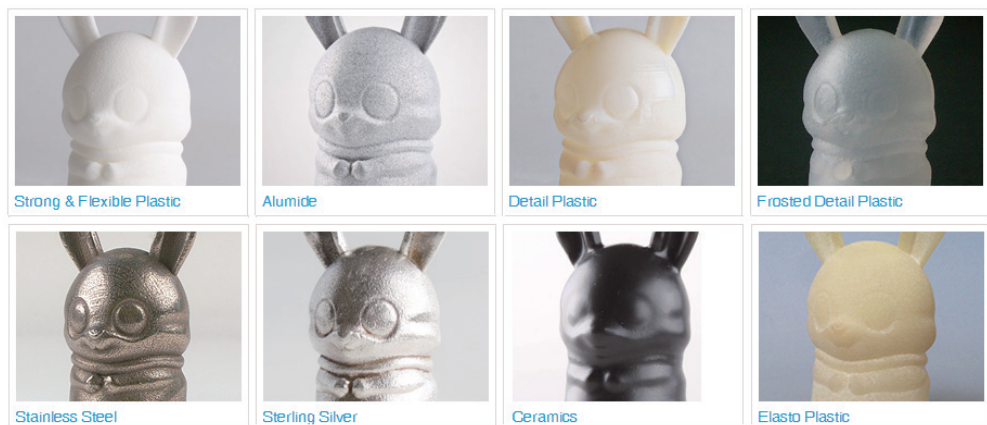


图 2-27 其他各种耗材的打印效果（图片来源：Shapeways）

金属类

Alumide（氧化铝）：一种闪耀的尼龙和铝粉的混合物。它适用于需要中等强度和细节度的产品，可以进行后期抛光使其更光滑，适于做镇纸和珠宝。“不防水，不可放入洗碗机，不能回收利用，对食品不安全，耐热度为 172°C ”。

Stainless（不锈钢）：通过向金属粉末床上层添加黏结剂层，再将产品用青铜浇铸浸渍来增加刚性，适于制作珠宝和结构件，可以被镀上铜和黄金。“防水，可放入洗碗机，不可回收再利用，对食品不安全，耐热度为 831°C ”。

Sterling Silver（纯银）：这种材料不是直接用于 3D 打印的，是先 3D 打印一个蜡模，再将

蜡模转换成石膏模具，再将熔化的银注入石膏模具成型。可后期机械或手工打磨，适于制作珠宝。

Titanium（钛）：采用直接金属激光烧结技术，将钛粉原料层层烧结成非常坚固且拥有高细节度的产品。可后期抛光，适于制作珠宝和功能件。

Gold（黄金）：工艺同 Silver（银），但原料在 14K 黄金中添加了一点铜来增加强度。适于制作珠宝。

Brass（黄铜）：铜和锌的混合材料，制作工艺和上面提过的银质物品一致。能够用聚氨酯类的油漆上色或镀金。适于制作珠宝、雕塑或为金银打印做测试。

Bronze（青铜）：和打印钢质物品类似，比较坚固。适合做装饰性的物品，比如雕像、钥匙和硬币等。

High Detailed Stainless Steel（高细节不锈钢）：和普通钢制作工艺相同，但使用 316L 不锈钢原料。在保持相同强度的基础上增加了更多的细节。适于制作珠宝、邮票模具和一些小的功能件。

Gold Plate（镀金）：不锈钢经 3D 打印之后镀金而成。

塑料和高聚合物类

BendLay：改进版 ABS，号称透明度高、柔性好。BendLay 可用在那些嫌 ABS 太硬，而又嫌柔性 PLA（soften PLA）不够紧实的地方。BendLay 不易卷翘，厂家更号称此线材对食品安全，可用在食物包装和医疗设备上。

LayBrick：是天然矿物填料（超细研磨白植土）和无毒聚酯的混合物，LayBrick 打印产品质感和石头类似。可用于制作建筑模型，可打磨，无须加热平台。

LayWoo-D3：再生木与高分子黏结剂结合的产物，可打印类似木材质地的一种线材。打印时卷翘度低，打印完成后可像对待木头一样对其进行钻孔，切割处理。所以它不论是看起来、摸起来、闻起来都很像真正的木头。

Strong and Flexible Plastic（坚固而有柔韧性的塑料）：此材料通过烧结工艺实现，它具有强度高、细节细致且柔韧的特性。它可以被染成多种颜色且可抛光处理。非常适于制作电话外壳、可穿戴产品和四轴飞行器框架。“不防水，不可放入洗碗机，不能回收再利用，对食品不安全，耐热度为 80℃”。

Detail Plastic（高细节塑料）：这是一种基于丙烯酸的感光聚合物，能制作具有高细节度，但对抗热和抗压要求不高的物品，非常适于制作小物件。“不防水，不可放入洗碗机，不能回收再利用，对食品不安全，耐热度为 48℃”。

Frosted Detail and Frosted Ultra Detail Plastic（磨砂细节和磨砂超细节塑料）：采用多点喷墨建模（Multijet Modeling）工艺，通过 UV 光固化将熔化的塑料通过多个喷嘴一层层沉积到一个平台上。这个工艺能够制作细节度非常高、特征丰富且很薄的壁，非常适合做微缩模型和其他模型。可后期上色，“不防水，不可放入洗碗机，不能回收再利用，对食品不安全，耐热度为 80℃”。

Elasto Plastic (弹性塑料): 这是一种还处于评估实验期的材料。它采用粉末烧结工艺, 非常地坚固且拥有超高弹力, 但细节度和平滑度较低, 非常适于制作手机壳和鞋子。它是一种易燃材料, “不防水, 不可放入洗碗机, 不能回收再利用, 对食品不安全, 耐热度为 90℃”。

Polyamide (聚酰胺): 采用粉末烧结技术, 这种材料相当结实而略有弹性。可以后期打磨、喷漆或染上各种颜色, 并可加入丝绒质感。非常适合打印雕塑、玩具和可穿戴产品。

Paintable Resin (可上色的树脂): 采用立体光刻技术, 可制作高细节度、需要中等机械抗性的产品, 可后期上色, 适合制作演示模型和雕塑。

Transparent Resin (透明树脂): 采用立体光刻技术制作全透明高细节度、非常光滑的产品, 可加入颜料, 适合制作演示模型和雕塑。

Prime Gray (灰色树脂): 采用立体光刻技术制作高细节度、非常光滑的灰色产品。非常适于制作产品原型、展示模型和桌面玩具等。

Rubber Like (类橡胶): 采用 EOS 的选择性激光烧结工艺, 由热塑性聚氨酯粉末床层层烧结获得, 此材料的特点是非常强壮且柔韧, 适于制作手机壳、夹子和所有需要在压力下有一点韧性的物品。该材料还处于测试阶段。

Durable Plastic (耐用塑料): 采用 3D Systems 公司的选择性激光烧结工艺将尼龙粉末熔融成坚固的产品, 适于制作原型和功能性零部件。

其他类

Food (食品): 有类似奶油那样的黏稠度的食材皆可用于打印, 比如巧克力、奶酪、糖等都可以通过注射器式的挤出喷头实现打印。

Ceramic (陶瓷): 这是 Shapeways 公司出品的第一款对食品安全的材料。它通过在陶瓷粉末床上层层添加黏结剂的方式来成型 (ZCorp 系列打印机), 之后产品再被烧制和上釉。它们并不太坚固, 细节度也不高, 但它们耐热, 适合做炊具和餐具。“防水, 不可放入洗碗机, 可回收再利用, 对食品安全, 耐热度为 500℃”。

2.4 金属3D打印技术大盘点

在 2.2.4 节中, 我们已提到 SLS 工艺用于金属打印。3D 打印金属是近年来发展很快的一个方向, 也是各国都乐意花大力气扶持的一个方向。目前, 不少金属 3D 打印机都已经能生产非常坚固的结构件、应力件和功能件, 成为 3D 打印家族中最有可能直接用于制造的技术, 应用也可遍及航空航天、医疗保健、仪器制造、电子行业等各个领域。下面就稍稍盘点一下那些神奇的金属 3D 打印机们。

虽然不同的厂商都为自己的金属打印工艺注册了各种商标名称, 叫法有所不同, 但其实原理是相似的, 都是逐层连续铺设金属粉末, 然后将其中的一些金属颗粒固定在一起, 以形成最终的物体, 原理如图 2-28 所示。

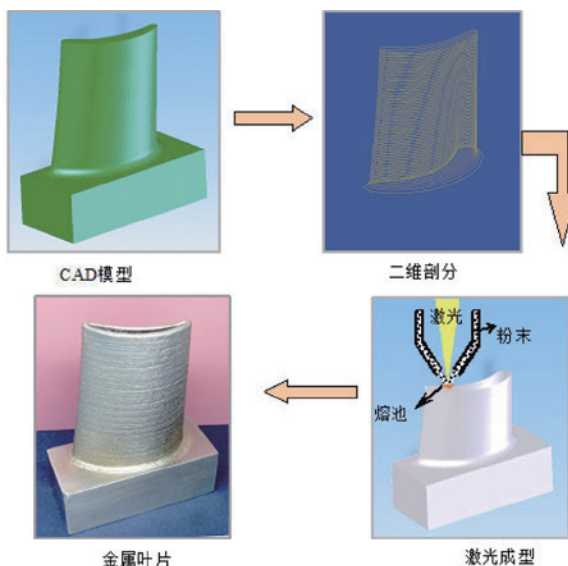


图 2-28 金属 3D 打印原理图（这里为激光近净成型工艺 LNSF）（图片来源：西安铂力特）

金属打印主要分为三大类：**激光烧结技术（Laser Sintering）**是使用激光束选择性加热粉末颗粒使其融合；而**黏结剂喷射（Binder Jetting）**金属打印技术的领军企业是 ExOne，黏结剂被选择性地喷射到多层的不锈钢、青铜、钨粉末上，然后将初步黏结而成的物体放置在熔炉中注入额外的熔融金属固化；但是，激光烧结或黏结剂喷射技术形成的物体并不是 100% 的致密，**电子束熔炼（EBM, Electron Beam Melting）**利用电子束选择性地融合粉末金属层解决了这个问题。下面，我们对各种金属 3D 打印技术进行详细介绍。

2.4.1 SLS、SLM 和 DMLS 技术

这 3 种技术只是专利名称和技术细节上有所不同，从原理上讲却都大同小异。大体上都是指将粉末状的材料（通常是金属材料）铺一层在工作台上，并将材料加温至略低于熔点，然后用高能激光束将金属粉末熔化并与上一层融合成一个实心整体，而未被扫描到的粉末材料仍呈粉状作为工件的支撑，一层扫描完成之后，工作台下降一个层高，再铺下一层粉末，重复上述过程一层层累积直至完成三维成型。

选择性激光烧结（SLS）技术

在 2.2.4 节，我们已介绍过 SLS。选择性激光烧结技术（SLS）使用激光束选择性加热粉末颗粒使其融合，精度为 0.1 ~ 0.2mm。具体来讲，SLS 采用的是一种金属材料与另一种低熔点材料（可以是低熔点金属或有机黏结材料）的混合物，在加工过程中，低熔点材料熔化或部分熔化，但熔点较高的金属材料并不熔化，而是被熔化或部分熔化的低熔点材料包覆黏结在一起。因此，形成的三维实体为类似粉末冶金烧结的坯件，实体存在一定比例的孔隙，不能达到 100% 密度，力学性能也较差，常常还需要经过高温重熔或渗金属填补孔隙等后处理才能使用。

SLS 技术由美国得克萨斯州立大学的 Carl Deckard 博士和 Joe Beaman 博士于 20 世纪 80 年代中期研发。后来他们俩成立了 DTM 公司专门研发 SLS 机器，2001 年 DTM 公司被 3D

Systems 公司收购。因此 SLS 技术的代表机型当之无愧的是 3D Systems 公司的 sPro 60SD、60HD、140 和 230 这 4 款机型，如图 2-29 所示。



图 2-29 3D Systems 公司的 SLS 打印机 sPro

选择性激光熔化（SLM）技术

SLM (Selective Laser Melting)，选择性激光熔化。SLM 是在选择性激光烧结（SLS）技术基础上发展起来的，但又区别于 SLS。SLS 工艺中粉体未发生完全熔化，成型件中含有未熔固相颗粒，直接导致孔隙率高、致密度低、拉伸强度差、表面粗糙度高等工艺缺陷。为获取全致密的激光成型件，同时也受益于 2000 年之后激光快速成型设备的长足进步（表现为先进高能光纤激光器的使用、铺粉精度的提高等），SLM 工艺迅速发展起来。相比于 SLS，**SLM 不依靠黏结剂**而是直接用激光束完全熔化粉体，成型性能得以显著提高。经 SLM 净成型的构件，成型精度高，综合力学性能优，可直接满足实际工程应用，在生物医学移植体制造领域具有重要的应用，如图 2-30 所示。

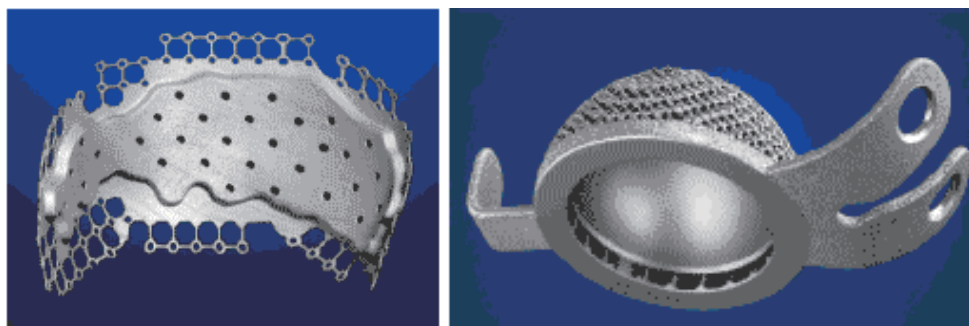


图 2-30 SLM 技术制造的钛合金头盖骨和关节窝生物移植体（图片来源：德国 Fraunhofer）

SLM 关键技术特点体现在如下几个方面。

- 直接制造高性能金属零件，省掉中间过渡环节；生产出的工件经抛光或简单表面处理可直接做模具、工件或医学金属植入体使用。

- 可得到冶金结合的金属实体，密度接近 100%；SLM 制造的工件有很高的拉伸强度。
- 由于 SLM 工艺采用的激光束光斑细小，产品具有很高的尺寸精度（精度可达 0.02mm）、较低的粗糙度，高于 SLS 的工艺水平。
- 适合各种复杂形状的工件，尤其适合内部有复杂异型结构（如空腔）、用传统方法无法制造的复杂工件。
- SLM 最大的问题在于熔化金属粉末时，零件内部易产生较大的应力，复杂结构需要添加支撑以抑制变形的产生。此外，零件性能的稳定控制较为困难。

SLM 技术由德国夫琅和费学院于 1995 年与当时的 F&S Stereolithographie-Technik 公司合作研发并申请获得相关专利。如今，SLM 技术的创始人 Dieter Schwarze 博士在 SLM Solutions 公司。SLM Solutions 公司出品的 SLM500 机型如图 2-31 所示。

3D Systems 公司也出品了采用 SLM 技术的金属 3D 打印机：sPro 125 和 250。3D Systems 公司称它们为直接金属选择性激光熔融 3D 打印机。它们能生产高精度、高复杂度的金属零件。打印层厚可达 20mm，可打印的金属包括钛、不锈钢、钴铬合金、工具钢等，所以能够应用在航空领域以及医疗保健领域（比如为整形外科、颌面修复和牙科治疗提供植入产品）等。



图 2-31 SLM Solutions 公司出品的 SLM500 机型

直接金属激光烧结（DMLS）技术

DMLS（Direct Metal Laser Sintering）技术由德国 EOS 公司开发，基本原理是 SLS 的进一步发展，**把热塑料黏结剂改为金属黏结剂就是 DMLS**。此外，DMLS 是边铺粉边烧结的，而 SLS 是先铺整层粉末，然后激光扫描烧结的。

在 DMLS 工艺中，由于粉末的颗粒度很细，最小叠层厚度仅为 0.02mm，因此制成的模具或零件的精度很高。这种方法制造的模具如果采用抛光处理，可以达到近似镜面的表面质量，成为高质量的模具，同时具有良好的机械性能，接近一般锻造构件。EOS 公司出品的 EOSINT M 系列机型非常类似 3D Systems 公司的 sPro 系列机型，能打印铝、钴铬合金、钛、镍合金和钢。

DMLS 的打印案例如图 2-32 所示。

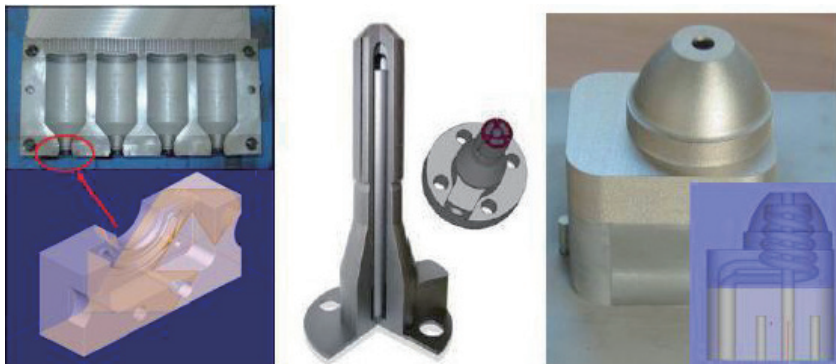


图 2-32 DMLS 的打印案例：随形冷却模具（图片来源：idnovo）

2.4.2 LENS/LNSF/LPF/DMD/LC/DLF：激光近净成型

这节要介绍的这项工艺被不同的厂商冠以不同的别称，如**激光工程化净成型**（LENS，Laser Engineered Net Shaping）、**激光近净成型**（LNSF）、**激光粉末成型**（LPF，Laser Powder Forming）、**直接金属沉积**（DMD，Direct Metal Deposition）、**激光固结**（LC，Laser Consolidation），此外还被称为**直接光制造**（DLF，Directed Light Fabrication）。与 SLM、DMLS 等工艺用激光照射**预先**铺展好的金属粉末不同，如图 2-33 所示，在 LENS 工艺中，激光照射喷嘴输送的粉末流，即激光与输送粉末**同时**工作，因此无须粉床。最早由西班牙 Sandia 国家实验室研发，利用激光束等高能束流熔化金属材料，在基体上形成熔池的同时将沉积材料（金属粉末或丝材）送入，随着熔池移动实现材料在基体上的沉积，工艺流程还可参见 2.4 节中的图 2-28。目前该工艺在国内使用比较多。

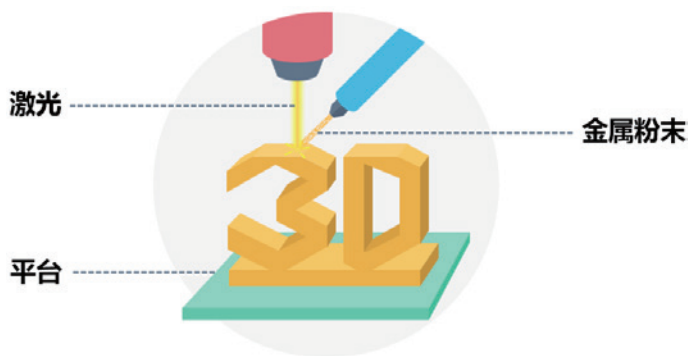


图 2-33 LENS 的技术原理（图片来源：thre3d.com）



提示：实际上，本节所介绍的工艺，连同 SLM 和 DMLS 等工艺，被统称为**激光熔覆成型**（LCF，Laser Cladding Forming）。这些技术的原理和加工方法基本相同，都是将快速原型制造技术和激光熔覆表面强化技术相结合，利用高能激光束在金属基体上形成熔池，将通过送粉装置和粉末喷嘴输送到熔池的金属粉末或事先预置于基体上的涂层熔化，快速凝固后与基体形成冶金结合。

LENS 可直接近净成型出全致密的金属零件或精坯。相比于 SLM 工艺，该工艺成型效率高，在直接制造航空航天、船舶、机械、动力等领域中大型复杂整体构件方面具有突出优势。但由于没有粉床的支撑功能，导致对复杂结构的成型较为困难，且成型精度略低。由于采用的激光光斑较粗，一般加工余量为 3 ~ 6mm。

LENS 技术除了制作新模型，同时也善于修补和在已有物件上二次添加新部件，因此应用面更加广泛。能打印不锈钢、殷钢和钛。采用 LENS 技术的代表机型是来自 Optomec 公司的 LENS 850R。LENS 打印机以及加工叶片的过程如图 2-34 所示。



图 2-34 LENS 打印机以及加工叶片的过程（图片来源：Optomec）

2.4.3 EBM：电子束熔炼

通过激光烧结或黏结剂喷射技术生产的金属物体都非常坚固，可以用于工业或其他方面的应用。但是，它们形成的物体并不是 100% 的致密。电子束熔炼（EBM，Electron Beam Melting）解决了这个潜在的问题，该技术与 DMLS、SLM 原理相似，只是采用的热源不是激光，而是在一个高度真空的打印腔中采用电子束来完成对金属粉末的熔融，如图 2-35 所示。通过高速电子轰击金属粉末，产生的动能转化成热能来熔化金属粉末。

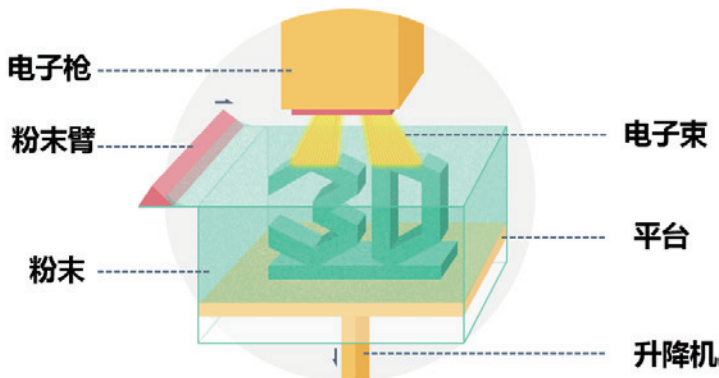


图 2-35 EBM 的技术原理（图片来源：thre3d.com）

由于打印过程在真空中进行,EBM 技术更适合打印那些易氧化或易和空气中某些元素进行反应的金属,比如钛。另外,EBM 采用纯净的合金粉末作为原材料,而不需要像 SLS、SLM 或 DMLS 那样在粉末中添加添加剂,因此也无须在打印后附加加热工序才能获得打印件机械特性。EBM 打印机以及打印案例如图 2-36 所示。



图 2-36 EBM 打印机以及打印案例 (图片来源 : Arcam)

EBM 最大的不足是设备需要严格的真空环境,电子束成本较高。另外,电子束聚斑效果较激光略差,导致零件的加工精度和表面质量略差(精度为 0.13 ~ 0.20mm),精度较选择性激光熔化(SLM)略低,但高于激光近净成型工艺(LNSF)。目前 EBM 工艺仅限于高价值的构建材料,包括各种钛和钴铬合金。这些材料目前主要用于航空航天及其他特殊的工业部门。

瑞典一家名为 Arcam 的公司使用 EBM 技术率先开发了 EBM 3D 打印机,使打印结果达到了非常高的品质,该技术在真空中逐层建立完全致密的金属物体。电子束快速成型速度快,是目前 3D 金属打印类中打印速度最快的,可达 15kg/h。2013 年 3 月, Arcam 公司推出了 Q10 型 3D 打印机,是专门为其公司的假肢市场生产的机型,用来替换之前的 A1 型 3D 打印机。另外, Arcam 公司还有专门针对航空市场的 A2 系列机型。

2.4.4 EBDM : 电子束直接制造

电子束直接制造(EBDM, Electron Beam Direct Manufacturing)技术是由美国 Sciaky 公司于 2009 年开发的一种新技术。与之前介绍的电子束熔融技术(EBM)不同, Sciaky 公司技术的独到之处在于:它将打印材料直接送进打印头,用电子束直接在机头熔融和打印材料,如图 2-37 所示。所以 EBDM 技术可以说是一滴一滴地打印金属物品的,其物品制作的精度和质量都非常高,更关键的是它基本不产生任何废料,节省了大量的原材料——考虑到金属的价格,这对降低成本有非常大的作用。

美国计划用 EBDM 来生产第五代隐形战斗机 F35 的多个零件,现在已经开始在做各种苛刻的检测。假设 F35 将生产 3 000 架,采用 EBDM 技术仅副翼这一个零件就能节省 1 亿英镑。很多钛合金零件都有希望采用 EBDM 技术,这项技术将是在不牺牲质量的情况下降低成本的关键。目前 EBDM 技术可以直接生产的金属包括钛、钽、钨镍合金等,能打印的最大尺寸为 19 英尺 × 4 英尺 × 4 英尺(约 5.7m × 1.2m × 1.2m)。去年 Sciaky 公司总共生产了 1 万 多斤金属制品,

其客户包括美国国防部、空军、洛克希勒、波音等。EBDM 的打印案例如图 2-38 所示。

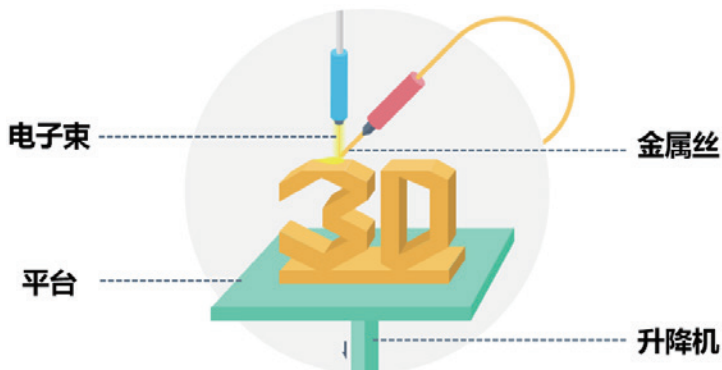


图 2-37 EBDM 的技术原理（图片来源：thre3d.com）



图 2-38 EBDM 的打印案例、EBDM 用于 F35 的副翼制造（图片来源：Sciaky）

2.4.5 金属 3D 打印技术小结

除了上面提到的金属 3D 打印技术，还有 UAM（Ultrasonic Additive Manufacturing，**超声波增材制造**，也被称为 Ultrasonic Consolidation、UC，精度 0.013mm，如图 2-39 左边所示，原理与 LOM 类似都属于片材分层加工）、IFF（Ion Fusion Formation，**离子熔化成型**，如图 2-39 右边所示，原理类似于 EBDM、LPF，造价相对便宜但速度相对较慢）等，在此就不详细做一一介绍了。

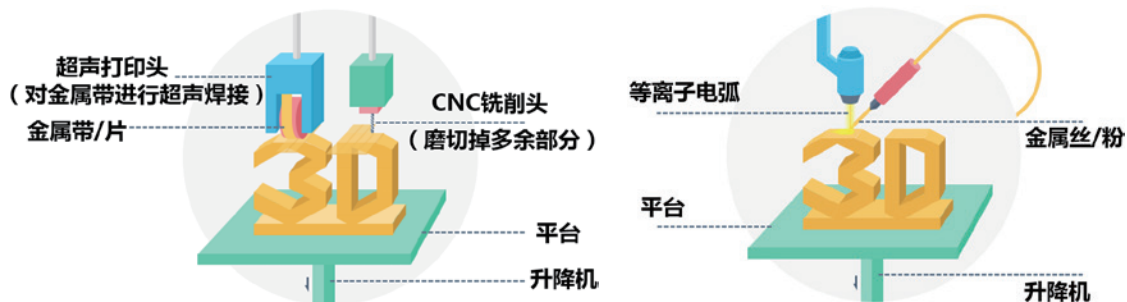


图 2-39 UAM（左图）和 IFF（右图）的技术原理（图片来源：thre3d.com）

最后，我们对金属 3D 打印进行一个小结。金属 3D 打印使得产品“直接制造”成为了可能，而不再局限于制造非功能性的模具。随着机械性能的不断提升，零件的致密性、强度已经与锻件基本相当，将来甚至还会有所超越。

然而，目前的金属 3D 打印构件都不能直接形成符合要求的零件表面，都必须经过进一步的机械加工，去除表面多余的不连续的不光滑的金属，才能作为最终使用的零件，因此，尽管 3D 打印可以获得复杂的空间结构和一些复杂的管路和腔体，但却无法对这些管路和腔体内部进行机械加工。因此，3D 打印虽可一步直接完成很多复杂零件的成型，但其还不具备直接取代传统机械加工的能力。

目前金属打印生成的零件表面精度一般在 0.1 ~ 5mm 之间。相比之下，目前市场销售的 2D 激光打印机点阵精度在 1 200dpi 左右，即 0.02mm，这个精度可以获得近似光滑的曲面。而要把金属 3D 打印精度提高到 0.1mm 以下还有很大困难，不过铺粉预处理、激光超快速熔化和凝固等技术的出现会为提高激光成型的精度提供很大的帮助。

此外，目前金属激光打印的速度还是较慢的，每小时打印重量大多数都在 1kg 以下，快一些的也不过 9kg/h 左右；若改用电子束直接制造技术（EBDM），最快也只有 20kg/h。要实现工业化生产，特别是大规模化生产，这个速度是不够的。现在的激光成型基本还是单光头单层铺粉作业，未来为了提高打印速度和应对超大型构件打印，可设计多光头多层铺粉同步打印。

2.5 两大阵营：工业级打印机与桌面级打印机

3D 打印机通常的划分是工业级和桌面级，桌面级一般也叫个人级、民用级、消费级。通常可加工超大尺寸的产品、价格昂贵的设备称之为工业级设备，一般使用 SLS、3DP 等技术，如 Objet 1000、Zprinter 系列的设备，主要应用于汽车、国防航空航天、工业机械、消费品、家电等工业领域。桌面级设备加工产品的尺寸一般较小，目前大部分使用的是 FDM 技术。工业级设备，如 Objet 1000，能打印 100cm×80cm×50cm 尺寸的成型产品，而桌面级设备的加工尺寸一般都在 20cm×20cm×20cm 左右。

2.5.1 工业级打印机：两个巨头的主战场

在 3D 打印领域，3D Systems 和 Stratasys，是两个不得不提的名字，它们争斗了近 30 年，持续上演着双雄争霸，它们的故事演绎着一个行业的发展轨迹。



说明：3D Systems 公司的技术优势和特色有：SLA（光固化立体成型）的鼻祖，全彩 3DP 打印。3D Systems 产品线涵盖个人级 3D 打印机（如 Cube、CubeX、3DTouch 系列等）、专业级 3D 打印机（如 ProJet 3500、ProJet 7000、Zprinter 650 等）和生产级 3D 打印机（如 SLS 工艺的 sPro 230，SLM 工艺的 sPro 250，最大的 SLA 打印机 iPro 9000 等）。

Stratasys 公司的技术优势和特色有：FDM（熔融沉积成型）的鼻祖，光敏固化技术精细度高、多种材料同时混合成型。Stratasys 的 3D 打印机产品线，根据可打印物品体积的不同，分为 MakerBot 系列、IDEA 创意系列（如 uPrint SE）、Design 设计系列（如 Objet 系列打印机）与 Production 产品系列（如 FOCUS 系列打印机）。

首先介绍 3D Systems 公司。创始人查尔斯·W·哈尔（Charles W. Hull）称得上是一位发明家，现年 70 多岁的他已经获得了 60 多项专利，其中最著名的当属为 3D 打印技术的普及铺平道路的“Stereolithography，立体平板印刷、立体光刻、光固化立体成型”技术。

1982 年，在紫外线设备生产商 UVP 担任副总裁的哈尔尝试将一种可被紫外光固化的液态感光树脂用于快速成型。哈尔的方法对于当时的快速成型工艺是一个巨大的突破。一方面，它加快了制造物品的速度，体积较小、构造相对简单的物品能在几小时内打印完成；另一方面，打印物品的体积也得到了提升，大多数光固化立体打印机能打印出 $50\text{cm} \times 50\text{cm} \times 60\text{cm}$ 的部件，有的机器则能打印出高达 2m 的物体。

1986 年 3 月，哈尔为这项技术申请了专利，随后他离开 UVP，成立了 3D Systems 公司，致力于将该技术商业化。为了让机器能够更加精确地将 CAD 模型打印成实物，哈尔又研发了著名的 STL 文件格式。STL 格式将 CAD 模型进行三角化处理，用许多散乱无序的三角形的小平面来表示三维物体，如今已是 CAD/CAM 系统接口文件格式的工业标准之一。3D Systems 的 3D 打印机如图 2-40 所示。



图 2-40 3D Systems 的 3D 打印机

不过，光固化立体成型技术也有自己的缺陷。由于采用紫外光对物体进行固化，这项技术所使用的材料有一定的局限，而且无论是机器本身还是光固化材料都价格高昂，这使得基于该技术的快速成型与 3D 打印技术的普及速度受到了限制。

与此同时，另一项同类技术的出现也对 3D Systems 公司造成了一定的冲击。20 世纪 80 年代中期，身为传感器制造商 IDEA 的联合创始人和销售副总裁的斯科特·克伦普（Scott Crump）决定设计一个能快速生产模型的机器。与光固化立体成型使用的光固化材料不同，克伦普的方案采用的材料是热塑性塑料。这一技术被克伦普命名为熔融沉积成型（Fused Deposition Modeling），并于 1989 年创立了 3D 打印机的制造商 Stratasys 公司，担任 CEO 至今。

1988 年，3D Systems 推出了第一台基于光固化立体成型技术的 3D 打印机。尽管体积庞大且售价高昂，但它的问世标志着 3D 打印商业化的起步。1990 年，3D Systems 从 UVP 公司购买了光固化立体成型的专利，3D 打印机的量产随之加快，公司也于当年在纳斯达克挂牌交易（2011 年转至纽交所）。

与此同时，Stratasys 也在克伦普的带领下快速成长，于 1992 年推出了第一台基于熔融沉积成型的 3D 打印机——“3D 造型者（3D Modeler）”。公司于 1994 年在纳斯达克上市，并先后推出了多款面向不同行业和市场的 3D 打印机，如图 2-41 所示。根据 Wohlers 2012 年的报告，2011 年 Stratasys 公司占有增材制造市场 41.5% 的市场份额，连续 10 年成为市场领头羊。



图 2-41 Stratasys 的 3D 打印机

目前，Stratasys 已成功打造了 uPrint、Dimension 和 Fortus 3 个品牌。其中桌面级的 uPrint 最低售价不到 16 000 美元，可打印中等大小的模型，支持彩色打印，适合工程师和教育工作者。落地式的 Dimension 问世于 2002 年，是目前全球市场最畅销的 3D 打印机系列之一，售价约 32 000 美元，使用坚固的 ABS 塑料作为打印材料，不仅能打印普通模型，还能打印车内组件，甚至是航空和医疗领域的零部件。Fortus 被称为“3D 生产系统”，覆盖从桌面级打印机到大型 3D 打印机在内的多款设备，打印物品的范围从概念模型到高要求的终端配件和制造工具，适用的领域则包括航空、汽车、商业、教育、医疗和军事等，当然售价也相对较高。Stratasys 还推出了廉价的桌面级专业 3D 打印机 Mojo 系列，售价低至 9 900 美元，能快速打印出较小的 ABS 模具。此外，2012 年 4 月，Stratasys 还合并了以色列公司 Objet，后者同样是 3D 打印机领域的巨头之一，其因参与《侏罗纪公园 3》、《钢铁侠 2》、《阿凡达》中的各尺寸模型的制作而在业界小有名气。2013 年 6 月 20 日，Stratasys 宣布收购日渐崛起的桌面级 3D 打印公司 MakerBot。两者的合并意味着以 Stratasys 为代表的专业级打印设备和以 MakerBot 为代表的消费机型在技术和品牌上的弥合，并开创面向学校和企业的中端市场的可能。

相对于 Stratasys，3D Systems 在面向个人和小工作室的入门级 3D 打印机市场的步伐要快得多。这家公司在 2009 年就推出了第一款 10 000 美元以内的 3D 打印机。2010 年 10 月和 2011 年 8 月，它又连续收购了两家廉价 3D 打印机生产商—Bits From Bytes（BFB）和 BotMill。两家公司的产品都以学生和个人爱好者为目标用户，目前 BFB 的 RapMan 系列最低售价为 1 390 美元，BotMill 旗下的 Axis 更是仅需 999 美元。3D Systems 又推出了面向家庭用户的迷你 3D 打印机，能打印 14cm×14cm×14cm 的物体，售价 1 299 美元起。除了个人和家庭市场，专业和生产级别的 3D 打印机也是 3D Systems 的主攻方向。由于材料的局限性，尽管光固化立体成型始终是公司的核心技术，3D Systems 也推出了基于选择性激光烧结（SLS）和金属激光烧结（DMLS）技术的产品。3D Systems 的产品支持多种材料，覆盖汽车、航空航天、消费电子、娱乐、医疗等多个领域。公司又以 1.37 亿美元的总价收购了专业 3D 打印机制造商 Z Corporation 和扫描设备生产商 Vidar。目前，3D Systems 已拥有超过 360 项美国专利，即便在 Stratasys 收购获得了 Objet 公司的 110 项专利后仍保持着领先。值得一提的是，3D Systems 公司还收购了 Geomagic

公司，这是一家全球领先的数字化解决方案供应商，包含数字化设计、数字化扫描、检测产品。

当然，除了 3D Systems 和 Stratasys 公司，还有其他公司也觊觎利润丰厚的工业级打印机市场。在两大巨头的压制之下，这些公司只能以产品的独特性来打入这个市场。以爱尔兰的 Mcor 公司为例，针对市面上绝大多数打印机不能打印全彩的现状，创新性地采用容易上色的纸质材料来进行彩色 3D 打印。所研发的 Matrix 300+ 彩色 3D 打印机，采用称之为选择性沉积层压（SDL，Selective Deposition Lamination）的工艺，其原理就是利用现有的纸张将其打成纸浆之后再逐层凝固上色，如图 2-42 所示。除了能实现彩色的 3D 打印，这种技术的好处有两点：一是耗材随处可见，只是普通 A4 纸；二是纸质的打印成品废弃后可以很容易降解，对环境非常友好，这比 3D Systems 的 Zprinter 全彩打印机使用石膏粉要环保。

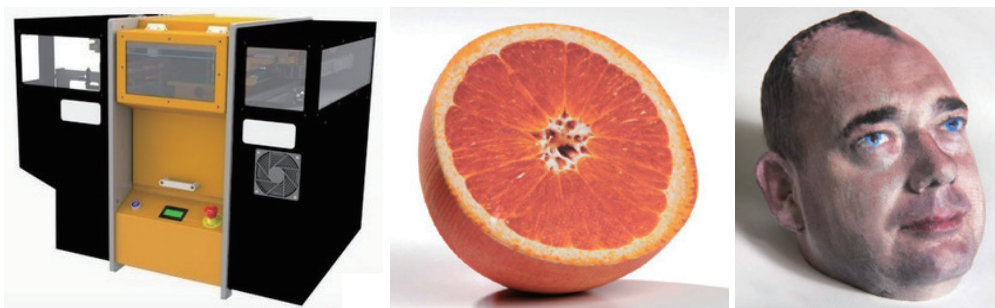


图 2-42 Mcor 公司的全彩 3D 打印机，以标准 A4 办公纸张作为打印材料



提示：Mcor 公司的 Matrix 300+ 彩色 3D 打印机采用的选择性沉积层压（SDL）技术与 2.2.8 节中介绍的 LOM（分层实体制造）技术原理相似且两者都采用纸张作为耗材。

在分辨率上，Matrix 可以达到 $5\,760 \times 1\,440 \times 508\text{dpi}$ ，据称这种打印机的打印速度要比传统的 3D 打印速度快 3 倍。

全球第三家 3D 打印上市公司 ExOne 的技术独到之处是可打印砂子，制作铸造用的砂模。旗下的 ExOne S-Max 也是最大的可打印金属、玻璃等材质的 3D 打印机，如图 2-43 所示。



图 2-43 ExOne 公司打印的螺旋桨砂模剖面（左），铸造的金属成品（右）

另一家公司 Voxeljet 的突出技术优势是产品非常庞大，最大可打印 $4\text{m} \times 2\text{m} \times 1\text{m}$ 的物体。Voxeljet 4000 是目前商品化能打印体积最大的 3D 打印机，如图 2-44 所示。

此外，德国 EOS 公司是激光 3D 金属打印的全球领导厂商，尤其在工业 3D 打印领域具备显著优势。DMLS 直接金属激光烧结技术就是它所研发的。



图 2-44 Voxeljet 4000 是最大的 3D 打印机 (图片来源 : Voxeljet)

上面介绍的这些工业级打印机因为体积庞大、价格昂贵，一般给人一种高深莫测的感觉。但实际上，操作起来其实是非常简单的。下面，我们就以 3D Systems 公司的产品为例，展示一下工业级打印的全过程。

我们采用的是 Zprinter 系列打印机，使用 3DP 工艺。原理和喷墨打印机一样。上一层粉，喷一层该截面形状的胶水，再继续上粉。可以选择透明和彩色两种胶水。

以下是打印机的外壳以及开盖后的图片，如图 2-45 所示。



图 2-45 Zprinter 工业级打印机外壳和开盖后照片

打印前，先将右槽的粉铲入左槽，慢慢降低左槽（同时升高右槽），再抹平表面。打印时左边槽升高、右边槽降低，一边把粉一层一层吹过去，一边上胶。打印的时间取决于产品体积大小。图 2-46 是打印完毕后用铲子挖开周边的样子。刚打印出来的产品很脆，很容易损坏。

挖得差不多了就用吸尘器一点点吸掉周边的粉末，如图 2-47 所示。吸尘器吸掉的粉末都回收在打印机下方，可以重复利用。

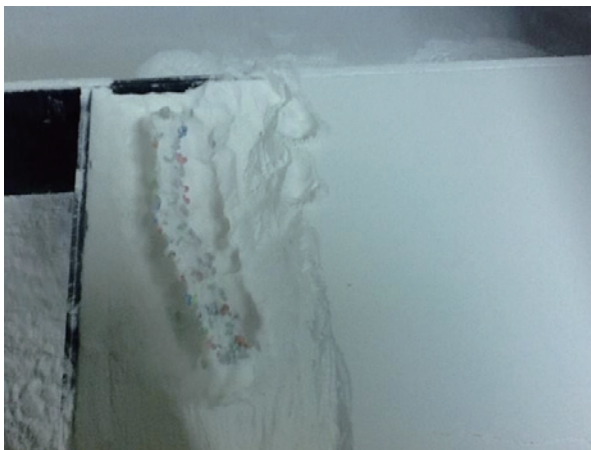


图 2-46 打印后用铲子挖开粉末



图 2-47 吸尘器吸干粉末

刨出来以后就可以把表面的粉末尽量吹掉了,吹干净以后上两次胶加固即可,如图 2-48 所示。
OK, 工业级 3D 打印是不是也非常简单?

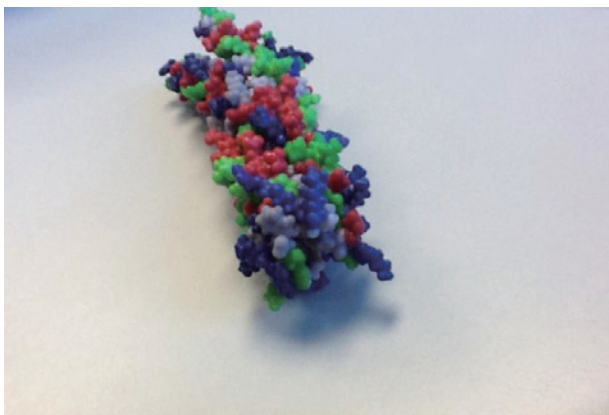


图 2-48 产品成型完成

2.5.2 桌面级打印机：创客们的多样世界

当下的 3D 打印机业界可以清晰地分为两类公司：一类是过去 30 年左右成立的，以生产价格在数万到数十万美元之间工业级设备为主的公司；而另一类则是从 2009 年开始崛起的桌面级打印机公司，生产的设备价格通常在几千美元左右。当然，大的工业级打印机生产公司（Stratasys 和 3D Systems）也已涉入桌面领域，推出了多款桌面级打印机。



提示：在最近几年里，桌面级打印机之所以能够兴起是因为当年工业级打印机的专利陆续到期失效。因此，MakerBot 和其他桌面级打印机公司们能够使用像 Stratasys 这样的工业先驱们的最初探索成果。

首先介绍一下美国 3D 打印机品牌。

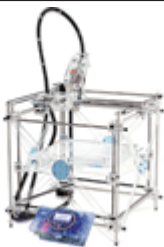
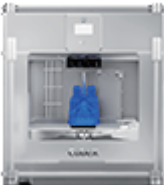

Stratasys 公司：

	产品系列	Mojo	产品型号	Mojo
	打印层厚	0.17mm	打印尺寸	127mm × 127mm × 127mm
	打印喷头	单头（单色）	官方报价	9 999 美元
	打印层厚	0.254mm	打印尺寸	203mm × 152mm × 152mm
	打印喷头	单头（单色）	官方报价	15 999 美元

MakerBot（也已属于 Stratasys 公司）：

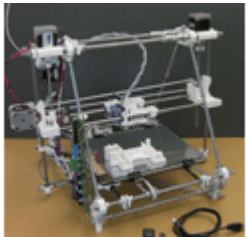
	产品系列	MakerBot	产品型号	MakerBot Replicator 2X
	打印层厚	0.1 ~ 0.34mm	打印尺寸	285mm × 153mm × 155mm
	打印喷头	双头（双色）	官方报价	2 799 美元
	打印层厚	0.1 ~ 0.34mm	打印尺寸	285mm × 153mm × 155mm
	打印喷头	单头（单色）	官方报价	2 199 美元

3D Systems 公司的桌面级打印机：

	产品系列	RapMan	产品型号	RapMan 3.2
	打印层厚	0.125mm	打印尺寸	270mm × 205mm × 210mm
	打印喷头	单 / 双可选	官方报价	1 030 欧元
	产品系列	Cubify	产品型号	CubeX
	打印层厚	0.125mm	打印尺寸	275mm × 265mm × 240mm (篮球大小)
	打印喷头	单 / 双 / 三可选	官方报价	2 799 美元
	产品系列	BFB (Bits from Bytes) 3DTouch	产品型号	3DTouch Single/Double/Triple
	打印层厚	0.125mm	打印尺寸	275mm × 275mm × 201mm
	打印喷头	单 / 双 / 三可选	官方报价	3 490 美元

欧洲 3D 打印机品牌列表。

RepRapPro：

	产品系列	RepRapPro	产品型号	RepRapPro Mendel
	打印层厚	0.1mm	打印尺寸	210mm × 190mm × 140mm
	打印喷头	三头（三色）	官方报价	1 089 美元

Ultimaker：

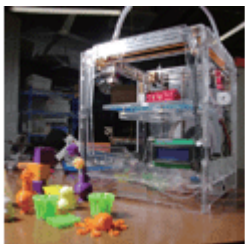
	产品系列	Ultimaker	产品型号	Ultimaker 2
	打印层厚	0.02 ~ 0.1mm	打印尺寸	230 × 225 × 205mm
	打印喷头	单头（单色）	官方报价	2 500 美元

最后再介绍几款中国国产 3D 打印机品牌。

太尔时代：

	产品系列	UP！系列	产品型号	UP！ Plus
	打印层厚	0.15 ~ 0.40mm	打印尺寸	140mm × 140mm × 135mm
	打印喷头	单头（单色）	官方报价	1 499 美元

DFRobot：

	产品品牌	DFROBOT	产品型号	Dream Maker
	打印层厚	≥ 0.05mm	打印尺寸	200mm × 200mm × 200mm
	打印喷头	单头（单色）	官方报价	6 899 元

AOD 公司（<http://www.aod3d.com>）：

	产品系列	AOD 智汇星	产品型号	X-Star 1.0
	打印层厚	≥ 0.04mm	打印尺寸	150mm × 150mm × 150mm
	打印喷头	单头（单色）	官方报价	3 999 元
	产品系列	AOD	产品型号	Artist 3.0
	打印层厚	≥ 0.04mm	打印尺寸	200mm × 200mm × 200mm
	打印喷头	单头（单色）PLA/ ABS/ 软性材料 / 石膏材料	官方报价	6 380 美元

（注意：以上的产品参数来自各厂商的宣传页面，并非在同一真实环境下进行对比获得。强烈建议读者在购买前，以亲自试打的效果为准。）

桌面级打印机价格低廉，性能指标也越来越向工业级打印机看齐。如果我们回顾第一台计算机庞大的身形，再放眼现在一部手掌大小手机的强大处理能力，就不难预测 3D 打印机的未来也将遵循当年计算机的发展足迹：越来越小型化，桌面级打印机也必将攻陷目前工业机的大部分疆土。

值得一提的是，桌面级打印机一般都是基于开源平台的，直接在共享设计资料的基础上做些改进就可推出自己的品牌，因此出现了百家争鸣的蓬勃发展局面。争奇斗艳、标新立异是桌面级打印机的一个最重要特色。下面就甄选几款有特色的桌面级打印机进行介绍。

全球首款桌面级多彩 3D 打印机 ProDesk3D

ProDesk3D 使用 PLA 或者 PVA 材料进行打印，精度误差控制在 25mm 以内，还能够实现打印 ABS 材料。这款设备最强大的地方在于，采用了 5 色的 PLA 墨盒系统，能够像传统的喷墨打印机一样通过双喷嘴对 3D 打印对象进行渲染。虽比不上 Zprinter 那样的全彩打印机，但也能形成多彩的颜色，如图 2-49 所示。



图 2-49 首款桌面级多彩 3D 打印机 ProDesk3D (图片来源 : botObjects)

可打印食品的 FAB@HOME

功能: 可用几乎任何材料进行打印，包括面粉、巧克力、蒙砂、黏土和橡胶，如图 2-50 所示。

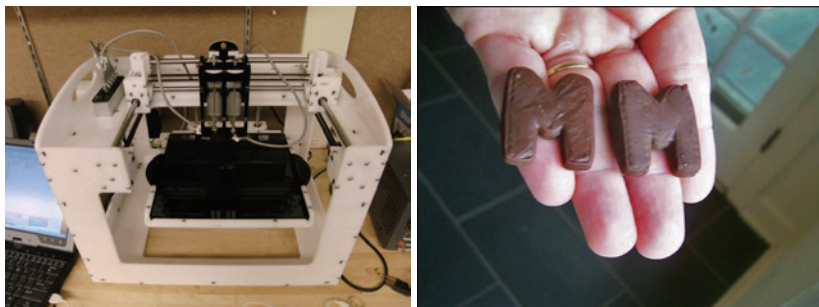


图 2-50 FAB@HOME 打印的巧克力食品

还有一款名叫 ChocoByte 的精巧机器，如图 2-51 所示，由悉尼的发明者 Quinn Karaitiana 设计，售价仅为 90 美元（约合人民币 559 元）。



图 2-51 一款名为 ChocoByte 的巧克力 3D 打印机

桌面上的专业级 3D 打印机 Form 1

一般而言，桌面级打印机的精度都不是太高。以颇受欢迎的桌面级 3D 打印机 MakerBot Replicator 2 为例，精度仅为 0.1mm。为了突破这一限制，Formlabs 推出的 Form 1 打印机的最高分辨率可以达到 0.025mm，意味着它已达到了工业级的精度。

就像大多数 3D 打印机一样，Form 1 也是通过逐层堆叠材料，但是不同于其他入门级的打印机，它使用液态光敏树脂进行光固化立体成型，这是一种可以呈现高精度细节的技术。

Form 1 打印机能够让你不需要花费更多的钱来获得更高品质的设备。这款产品使用液态光敏树脂进行光固化立体成型来实现精准打印，最大打印尺寸为 4.9ft×4.9ft×6.5ft。这款高精度 3D 打印机目前的市场售价仅为 3 299 美元，如图 2-52、图 2-53 所示。



提示：虽然 Form 1 打印机不贵，但使用的材料光敏树脂相比于廉价的 ABS、PLA 耗材来说却是比较昂贵的。

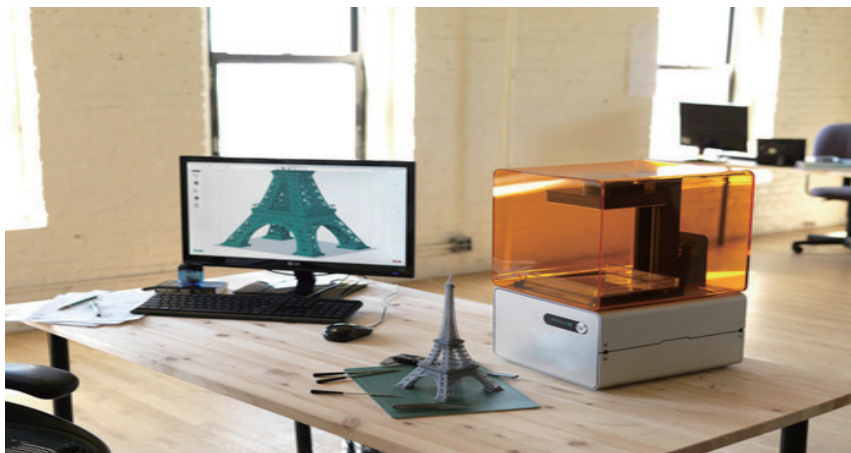


图 2-52 Form 1 打印机以及打印的埃菲尔铁塔（图片来源：Formlabs）

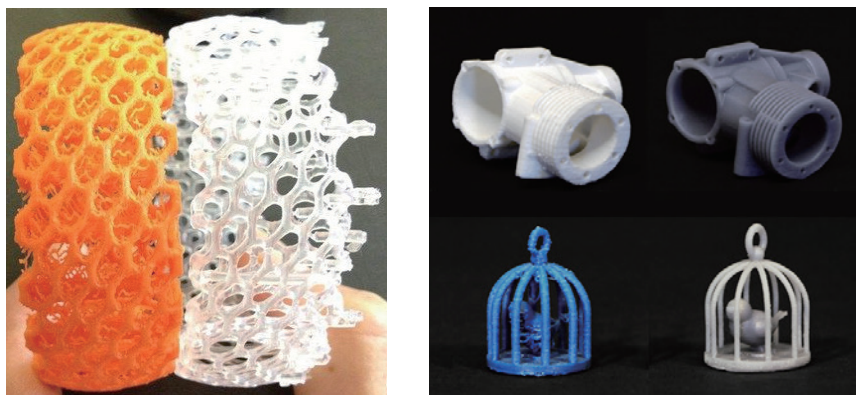


图 2-53 Form1 打印案例（右）与 FDM 桌面级 3D 打印机 MakerBot 打印案例（左）的效果对比

除了 Form 1，现在还有 B9Creator、Sedgwick、MiiCraft 等品牌的高精度桌面机，也都是用液态光敏树脂进行光固化立体成型的。

桌上的碳纤维 3D 打印机 Mark One

碳纤维这种材料强度是钢的 2 倍，重量却只有其三分之一。美国 MarkForg3D 公司推出一款造价仅为 4 999 美元的个人 3D 打印机 Mark One，如图 2-54 所示，可利用碳纤维进行桌面打印，产品既轻又硬，硬度是 ABS 耗材的 20 倍，甚至超过了铝合金。这款 3D 打印机精度为 $100\mu\text{m}$ ，成型尺寸为 $305\times 160\times 160\text{mm}$ 。



图 2-54 世界上首款碳纤维 3D 桌面级打印机 Mark One 及打印样件

最后再介绍一款奇特的 3D 打印机。一般情况下，3D 打印机需要放到一个水平的地方再进行作业。而一款名叫“Mataerial”的 3D 打印机器人打破了这种局限，它完全可以在垂直、光滑，甚至凹凸不平的表面打印产品，不需要任何支撑工具。机器使用热固性聚合物取代一般 3D 打印机常用的热塑性塑料，从喷头挤出来的材料能够瞬间凝固，并随着机械臂的摆动而调整出各种形状，如图 2-55 所示。

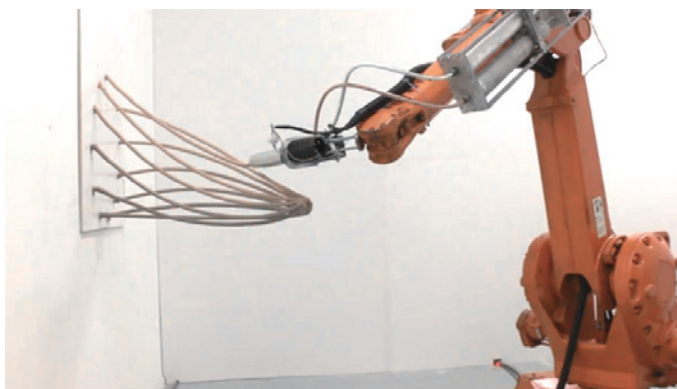


图 2-55 可在垂直墙面打印的 3D 打印机“Mataerial”（图片来源：Petr Novikov 和 Sasa Jokic）

2.6 3D打印与传统手办模型制作

目前受 3D 打印冲击最大的应该就是手办模型行业了，这个行业在未来很有可能会迅速消亡。手办模型属于典型的个性定制化、高附加值行业，一般都是手工生产限量版的 3D 模型，比如为

某部热映的好莱坞科幻大片定制里面的主角卡通模型，如图 2-56 所示。

传统的手办模型制作速度慢、成本高。精度虽然不错，但对于细节特别多的形状则无能为力，比如要在形状上刻上几千个花纹，对于手工几乎是不可能的任务。此外，由于事先没有使用 CAD 软件进行数字化设计，所以每一个模型都只能是唯一的，难以大批量复制。在做出最终模型之前，无法预览效果的好坏。如果最终成型后的效果不好则会遭到客户或项目甲方的否定，但为时已晚。



图 2-56 传统泡沫纸板制作的模型 VS. 用 ZCorp 3D 打印机做出的楼房模型
(图片来源 : 3D Systems)

而使用 3D 打印技术，可在项目早期阶段就通过数字化 3D 设计让客户预览到最终的效果，可增强设计者与客户之间的交流。3D 打印可生成具有复杂内嵌细节的建筑结构，也可制作出任意比例的模型。全彩色打印会给客户留下更深刻的印象并更有说服力。而且，因为有了数字化设计版本，所以可以简单、快速、低花费地复制相同的模型。模型可在设计的任何阶段打印输出，直接应用到城市规划布局上，如图 2-57 所示。

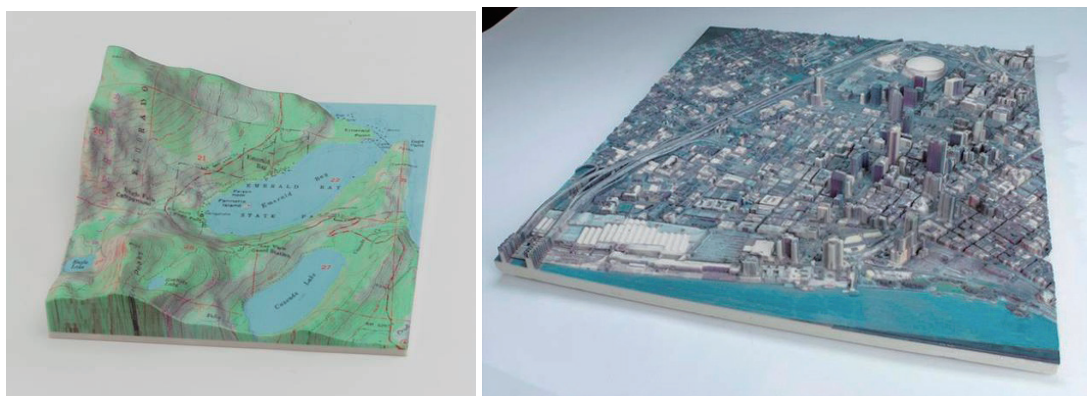


图 2-57 3D 打印全彩模型用于城市规划布局

2.7 3D打印机购买指南

购买 3D 打印机时，要根据实际的应用需求以及打印机的性能指标来进行选择。具体来讲，可重点考虑以下这些具体的性能属性。

最小细节分辨率

分辨率是 3D 打印机最重要的指标，因为它直接决定了输出的精度，同时它又是最令人困惑的指标之一：分辨率可能写成 Z 轴层厚、XY 轴精度、每英寸点数（dpi）、像素尺寸、束斑大小和喷嘴直径等。尽管这些参数有助于比较同一类 3D 打印机的分辨率，但是很难用来比较不同的 3D 打印技术。一般而言，**最重要的一项参数是 Z 轴层的厚度**，例如，MakerBot Replicator 2/2X 的打印层厚为 0.1mm。当然，不要轻易相信一些厂商宣称的精度，最好的办法是亲自试打一个具有微小细节的 3D 模型，比如细长的角、纤细的手指头等，然后仔细查看锋利的边缘和拐角清晰度、最小细节尺寸、侧壁质量和表面光滑度。

还有一个很重要但常常被忽略的参数，那就是**线宽（Thread Width）**。大部分的打印机拥有直径是 0.4mm 或是 0.5mm 的喷头，而线宽是由打印机喷头的直径来决定的（详情见 2.2.2 节），通常会大于喷头直径。事实上，3D 打印机画出来的圆，大小都会是线宽的两倍。举例来说：一个 0.4mm 的喷头画出来的圆最小直径是 0.8mm，而 0.5mm 的喷头画出来的最小直径则是 1mm，也即“**你能创造的最小物件不会小于线宽的两倍**”。

目前有两位美国的创客开发了用于评测 3D 打印机性能的测试片套装，用于测试控制平滑表面的能力、各种倒凹型和细节保留度、打印伸缩结构的能力等。当这些测试片被打印出来之后，你就可以测量这些部分的实际大小和角度，并与其 3D 文件相比较，从而判断打印机的各项性能。另外，Thingiverse 网站也提供了一个测试套件：The Essential Calibration Set（基本标定集），下载详见：<http://www.thingiverse.com/thing:5573>。

打印速度

根据 3D 打印机制造商所采用的 3D 打印技术的不同，3D 打印速度的评价标准也是不同的。有些用来指示在 Z 轴方向上打印一定高度所需的时间，通常用 inch/h、mm/h 来表示。那些具备稳定的垂直打印速度的 3D 打印机普遍采用这一技术参数，其不受被打印物体的结构复杂度和单次打印部件数量的影响。

另外一种评价打印速度的指标是打印一定体积所需的时间。一些 3D 打印机可以快速打印单个、结构简单的物体，一般采用这种指标。但是这种类型的 3D 打印机在遇到打印数量增加或者结构比较复杂的打印任务时，打印速度就会明显下降，因此不适用于打印速度要求较高的手板模型。

对于目前的 FDM 桌面级 3D 打印机而言，如果速度是你的首选，可考虑选择 Ultimaker。

3D 打印成本 / 部件成本

部件成本通常表示为每单位体积的成本，如每立方英寸的成本或每立方厘米的成本。

部件成本取决于所消耗的材料总量以及材料的价格。除了廉价的 ABS 和 PLA 塑料，石膏粉

也是非常便宜的。而且，未使用的石膏粉末可放入打印机中回收和再利用，因此成本可以达到其他 3D 打印技术的 1/3 到 1/2。

有些供应商提供的报价只包含了成型件所需模型材料的费用，而不含支撑材料或者 3D 打印过程中产生的其他损耗的费用，这种报价不是真实的、最终的报价。

材料属性

了解预期的应用和所需材料的特性，对于选择 3D 打印机来说很重要。每种技术各有所长也各有所短。

对于概念建模应用来说，实际的物理特性可能没有部件成本和模型外观那么重要。概念模型主要用于预览效果，可能使用后很快就被丢弃。而验证模型可能需要模拟最终产品的效果，需要实现与最终生产材料接近的功能特征，如具有可铸性或耐高温，一般需要在较长的时间内保持牢固。

每种 3D 打印技术都受限于具体的材料类型。对于个人 3D 打印，材料大致可分为非塑料、塑料、蜡这几类。非塑料材料常使用石膏粉与可打印的黏结剂，部件成品紧密而坚硬，可以通过浸润变得非常牢固。石膏粉结合独特的全彩色打印能力，可以制造出逼真的视觉模型，而无须额外的绘画或后期处理。但石膏强度不高，脆弱易碎。

塑料材料可以柔软、可以坚硬，有些还具有高耐温性。透明塑料材料、生物相容性塑料材料、可铸性塑料材料均有销售。有些 3D 打印的塑料成品是防水的，而有些却是多孔的，会因吸收水分导致产品膨胀而改变尺寸。

色彩

目前大多数打印机都只能打印单色。少数打印机加装多喷头（双喷头、三喷头）后，可打印双色、三色模型，如图 2-58 所示。



图 2-58 三色（三喷头）打印案例（图片来源：Cubify）

目前只有极少数几种打印机支持全彩色，如 3D Systems 公司的 Zprinter 3D 打印机，它可以达到几乎与 3D 打印模型一致的颜色，有 390 000 种颜色组合。

产品特点

有些 3D 打印机出奇制胜，在某项性能上有特别的优势。比如，CubeX 的打印尺寸足以容纳一个篮球，大的打印尺寸是它在同期产品中的亮点之一。

MakerBot Replicator 2X 所用的新型双喷头设计源自开源社区成员的方案，可以实现 2 个喷头同时打印 2 个不同参数的物体，这也是 Replicator 2X 在同期产品中的亮点之一。

选择专业的 3D 打印机销售网站

一般来说，选购商品时人们会第一时间想到去淘宝和京东类的网站，其实更推荐去专业的 3D 打印机网站进行咨询。这些专业网站往往由发烧友或业内人士建立，所推荐的机器一般性价比确实不错，此外 3D 打印机毕竟是个有较高专业技术的产品，目前还很难做到傻瓜式一键操作。专业网站的客服一般对机器技术和性能参数方面比较了解，在出故障时可以很好地指导你进行机器调试。特别是，目前国外的品牌暂时还不好返厂维修，售后链没覆盖到各个城市，因此选择一家专业的国内代理商尤其重要。

如果想自己从官方买的话，可以直接登录国外官网，并选购要买的打印机和耗材，然后去结算。但是这些公司都不会帮你缴纳关税，这个问题需要你自己解决。考虑到关税和运费，运到大陆地区的最终花费一般是国外官网报价的 2 倍左右。

总之，不同的 3D 打印应用会有不同的需求，明确自己的真正需求是选择 3D 打印机的关键。

第3章

剖析3D打印机：轮子是怎样发明的

《易经·系辞》有云：“形而上者谓之道，形而下者谓之器”。这里，“形而上”即指无形之道（抽象道理、原理规律），“形而下”即指有形之器（具体事物、设备器具）。在第2章中，我们已经介绍了关于3D打印的原理方法，也即形而上之道。然而，光凭焚香论道、席地侃谈是无法真正了解3D打印实际过程的。因此，道必须化为器，也即实实在在的设备，才能把3D打印真正运作起来。

在本章中，我们将手把手地教会你亲手组装一台3D打印机！这里用到了“庖丁解牛”的逆过程，把各个功能模块逐一组装好，然后再搭建成完整的一台机器。虽然在日常应用中，我们一般都不必自己组装设备，直接购买现成的打印机即可。国外也常提到一句著名的谚语：“不必重新发明轮子”（Don't reinvent the wheel）。然而在本章中，为了让读者对3D打印机有一个彻头彻尾的真正了解，我们还是非常愿意不辞辛苦地、手把手地教会你如何从无到有地发明轮子！

3.1 RepRap：开源3D打印机的鼻祖和奠基石

虽然3D打印机技术在将近30年前就已发明，但是其在最近几年能够这么“火”，很大程度上归功于桌面级3D打印机的迅速流行，如MakerBot Replicator 2在2012年赚足了全世界的眼球。作为桌面级3D打印机的祖师爷，RepRap（Replicating Rapid Prototyper，快速自我复制原型机）绝对功不可没。实际上，MakerBot、Ultimaker等绝大多数桌面级3D打印机就是基于完全开源的RepRap才发展起来的。

RepRap是一个3D打印原型机，它具有一定程度的自我复制能力，能够打印出大部分自身的（塑料）组件。该原型机从软件到硬件等各种资料均是免费和开源的（均在自由软件协议GNU通用公共许可证GPL之下发布），由此引发了全世界的创客（Maker）纷纷加入到3D打印机制作的狂潮。RepRap基于流行的开放源码的Arduino硬件平台，当前版本使用的是Arduino的衍生版本Sanguino主板。

RepRap项目由英国巴斯大学高级讲师Adrian Bowyer博士创建于2005年。至目前为止，RepRap项目已经发布了4个版本的3D立体打印机：2007年3月发布的“达尔文”（Darwin），

2009年10月发布的“孟德尔”(Mendel),以及在2010年发布的“Prusa Mendel”和“赫胥黎”(Huxley)。可以看出,开发者喜欢采用著名生物学家们的名字来命名,因为“RepRap就是复制和进化”。

由于机器具有自我复制能力,如图3-1所示,发明者设想可以廉价地传播RepRap给个人和社区,使他们在家里就能够制造复杂的设备和产品,而不需要昂贵的工业设施。而且,在此过程中RepRap是可以不断进化的,同时它的数量可以按指数成倍地复制增加。

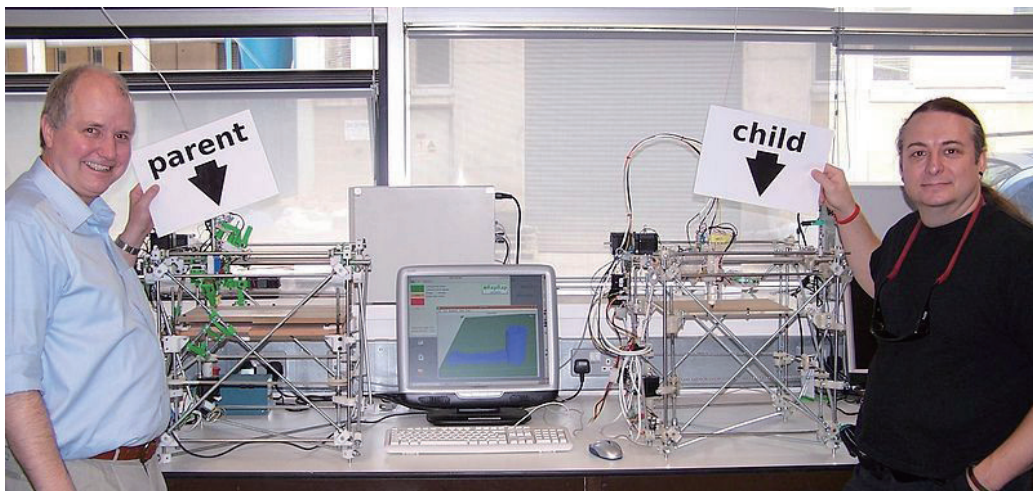


图 3-1 能够自我复制的 3D 打印机 RepRap, 右边的打印机由左边的打印机复制而出
(图片来源: RepRap.org)

3.2 MakerBot与Ultimaker: 桌面双雄

在桌面机领域,目前最具代表性的就是Ultimaker与MakerBot了。其中MakerBot“问道于江湖”更早,现在几乎“天下谁人不识君”,只是非常可惜的是,MakerBot商业化后已经不再开源,最终被合并到了工业巨头Stratasys的麾下。Ultimaker为后起之秀,在打印速度和精度上不断改进,且保持了开源精神,因此在本章3.3节中我们采用了Ultimaker为例介绍了3D个人打印机的组装全过程。

MakerBot

MakerBot是一家位于美国布鲁克林的创业公司,他们发明的MakerBot目标是家用型的、简易的3D打印机。MakerBot是FDM型打印机,使用ABS或者PLA塑料作为原料,如图3-2所示。最初的产品CupcakeCNC及Thing-O-Matic都是发轫于RepRap开源项目并且都是开源的。但是由于公司策略发生变化,新产品Replicator、Replicator 2/2X等都已经不再开源了(也由此导致了公司3位创始人中的2位已经先后离开),这对于推动3D打印机的进步是件非常可惜的事情。

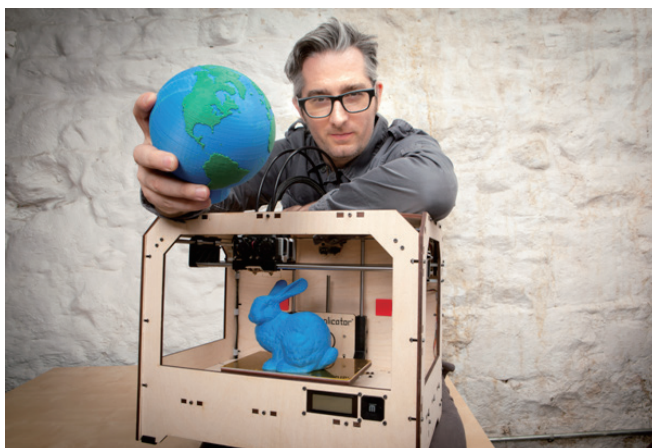


图 3-2 MakerBot CEO 布雷·佩蒂斯 (Bre Pettis) 以及 Replicator 打印机 (图片来源: MakerBot)

MakerBot Replicator 2 被视为稳定性最好的一款家用打印机, 售价为 2 199 美元。与第一代 Replicator 相比, 木质外壳变成了金属外壳, 里面有多层塑胶保护正在打印的物体。Replicator 2 的分辨率为 $100\mu\text{m}$ (Microns), 即 0.1mm, 每一层横切面都像一张纸一样薄, 所以打印出来的物体比较细致。MakerBot 还开发了自己的软件 MakerWare, 用来在打印前进行各项参数设定以及模型的可视化预览。Replicator 还装有 LED 灯, 除了有指示进程的功能, 打印过程在蓝色 LED 灯的照射下, 也显得科技酷感十足。

Ultimaker

Ultimaker 是由 3 位来自荷兰的年轻创客 (Maker) 共同开发的, 如图 3-3 所示。相比于 MakerBot, Ultimaker 具有更快的速度, 更高的性价比, 可打印更大的尺寸, 同时还是一款完全开源的 3D 打印机。Ultimaker 首次露面便受到了极大的好评。援引自《Make》杂志上的一句话: “这是对 3D 打印机的一大改进!”



图 3-3 Ultimaker 的 3 位创始人, 从左到右: Siert Wijnia、Martijn Elserman 和 Erik de Bruijn (图片来源: Ultimaker)

Ultimaker 自称是最快、最精准的 3D 打印机，采用 PLA 为耗材的 Ultimaker 分辨率最高可达 12.5 μ m，售价为 1 194 欧元。和 MakerBot 一样，Ultimaker 也是使用 ABS 塑料或 PLA 塑料来制作产品的，同属于 FDM 型打印机。Ultimaker 和 MakerBot 的不同之处在于，MakerBot 的马达安装在可动零件上，而 Ultimaker 的马达安装在打印机的框架上，因此 Ultimaker 打印机可以得到更好的稳定性以及更大的打印尺寸。此外，MakerBot 是依靠平台的移动来进行打印的，而 Ultimaker 则是依赖喷头的移动。相比之下，Ultimaker 的喷头更为精巧且重量很轻，因此打印速度快了几倍。在 Ultimaker 官方网站上甚至还有喷头移动速度为 350mm/s，挤出速度为 300mm/s 的演示视频。第二代产品 Ultimaker 2 在打印时只产生 49 分贝的噪声（差不多只相当于耳语声的 3 倍），比普通桌面 3D 打印机要安静得多。

除了 MakerBot 和 Ultimaker，越来越多的桌面机层出不穷，详情可参考第 2 章的 2.5.2 节“桌面级打印机：创客们的多样世界”。相信未来这些桌面级 3D 打印机会不断改进，向昂贵的价格和技术垄断“说不”，快速渗透到我们生活的每个角落。

3.3 Ultimaker 组装实战

Ultimaker 基于开源的 RepRap 打印机 DIY 平台，你可以根据零件清单，自己采购组装一款高速、大容积、高分辨率的开源 3D 打印机。

Ultimaker 有以下几个关键特点。

- 容易组装。
- 基于 FDM 成型工艺，且为准产品级别的设计！
- 双杆固定。所有的马达都非常稳固，移动时抖动非常小，最大可以达到 500mm/s 的电机移动速度。
- 超大打印容积。最大打印容积可达 230mm×225mm×205mm。
- 高分辨率。在每个方向的打印可达到 20mm 的精度。
- 开源，真正意义上的开源！不像某些所谓的开源打印机，打着开源的幌子，你根本无法找到他们的零件清单和相关软硬件源码图纸。如果你能力强，完全可以 DIY 一台 Ultimaker，升级改进它，然后推出自己的品牌（如 China Dreammaker，中国梦牌 3D 打印机）！

综上所述，Ultimaker 是一款性能相当优异，性价比也超级高的准工业级打印机。《Make》杂志在一次评比中，将“最佳开源硬件”、“最快速的”和“最精确的”这 3 项头衔给了 Ultimaker。

下面我们就详细介绍如何自己组装一台 3D 打印机。先把打印机的工作原理再捋顺一下：通过 X、Y 轴伺服电机带动打印头，从送料机送入一根热塑性细丝材料（ABS 或 PLA 塑料），经过打印头加热后，再用挤出头把熔融物挤出成一层薄片，这一层薄片在载物平台上冷却后迅速变硬。然后 Z 轴电机将打印头略微向上移动一层，再继续挤出堆积到上一层薄片，就这样反复操作，层层堆积，几个小时后，一个 3D 实体模型就“打印”堆积出来了。

3.3.1 Ultimaker 新到货开箱照

Ultimaker 组装套件开箱后的照片如图 3-4 所示。要将它们组装成一台可工作的 3D 打印机，一般需要 6 ~ 20 个小时。如果你想成为一名真正的创客（Maker），那就全部自己动手吧！这样才能获得对一台 3D 打印机真正的理解！



图 3-4 Ultimaker 新到货开箱照（图片来源：3djoy.cn）

组装完成后的效果如图 3-5 所示，是不是会让你有一种很有成就的感觉呢？

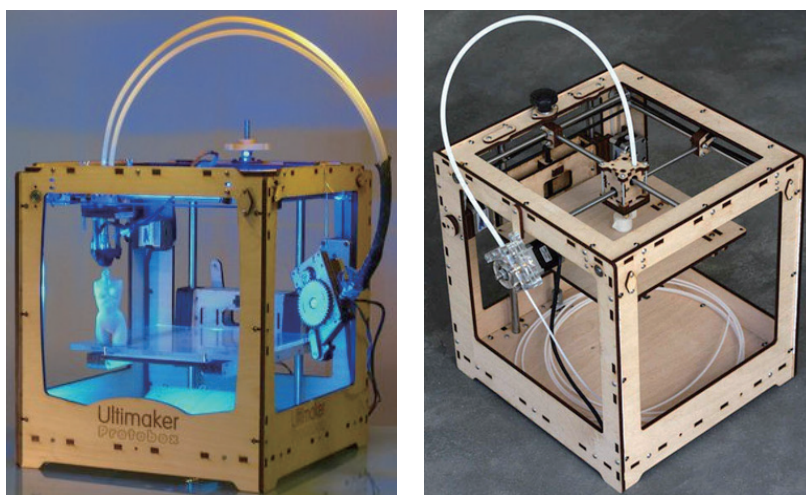


图 3-5 组装完成后的效果图（图片来源：3ders.org）

值得注意的是，各个型号的 3D 打印机组装过程并不相同，甚至同一型号的打印机也有多个版本。因此本节给出了整体的流程步骤，更多细节请参考厂商提供的用户手册和网络教程^{[47]、[55]}。

3.3.2 搭建框架

首先我们将 3D 打印机的框架（Frame，比如矩形盒式结构、矩形杆式结构、三角形结构、三角爪式结构等；当然你也可把框架类比为计算机的机箱，虽然这个类比不是很准确）搭建好，完成后的效果如图 3-6 所示。

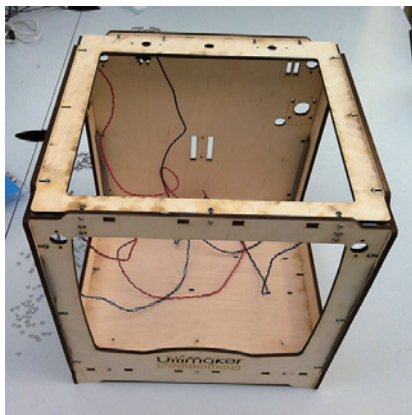


图 3-6 框架搭建完成的效果图（图片来源：Ultimaker wiki）

第 1 步：限位开关

限位开关的主要作用是：打印时，如果传输架已抵达机器的边缘，限位开关可确保机器立即停止移动。

1. 在前面板的背面上,使用 2 个 M3 螺栓(10mm)安装 2 个蓝色的限位开关,控制弹片朝上。
2. 在左面板的背面上, 安装 2 个红色的限位开关, 控制弹片朝下。
3. 在后面板的背面上, 安装 2 个黑色的限位开关, 彼此相对且弹片都朝左。在底部面板使用 12mm 的 M3 螺栓。对于后面板上方的限位开关, 在螺杆头部和木材之间使用 2 个垫片, 以便最后再做细小的调整。

完成后, 3 个面板的效果如图 3-7 所示。

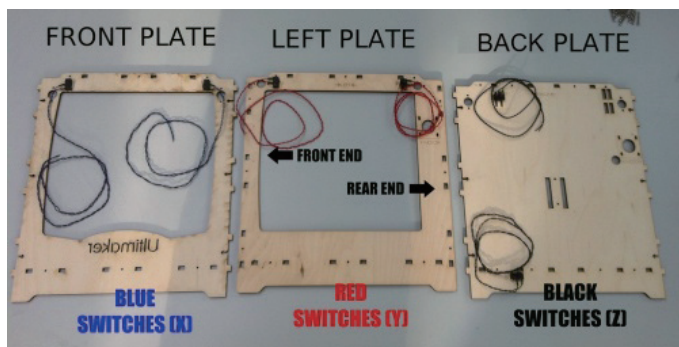


图 3-7 安装限位开关的面板

第 2 步：组装立体框架

1. 将后面板上有标记的一面朝向你。
2. 插上顶部面板，将顶部面板的插销嵌入后面板的插槽。
3. 同理，组装底面板和前面板。然后，轻轻将框架侧倒，如图 3-8 所示。请小心：此时还没有通过螺栓连接在一起！

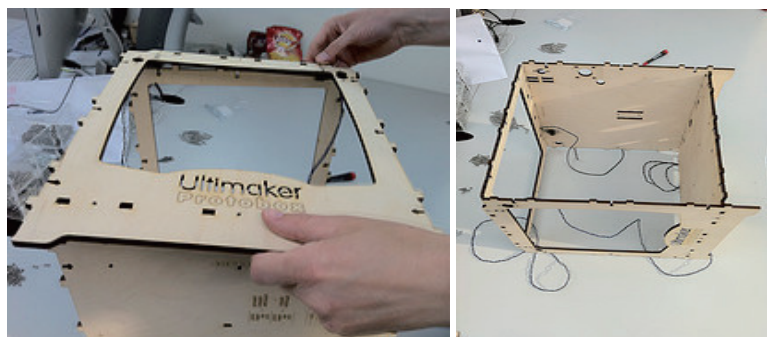


图 3-8 组装底面板和前面板，将框架侧倒

4. 从电缆包中取出 4 个电缆导管：2 个长的、2 个短的。

5. 将每个电缆导管对折，放置在标贴上。然后用蓝色胶带固定电缆导管，以保持它折叠，如图 3-9 所示。

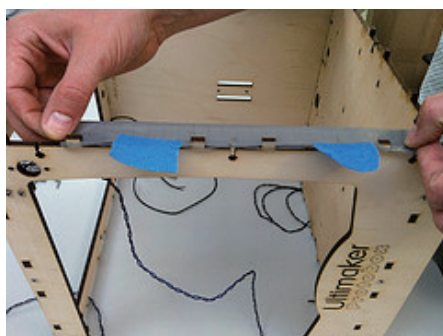


图 3-9 用蓝色胶带固定电缆导管

6. 用螺栓固定整个框架，至此框架搭建完成。

第 3 步：安装各种部件

1. 用自锁螺帽安装转轴支架固定板。

2. 在机器的底面（在后面），使用 2 个 16mm 螺栓安装 2 个标有 3A 的部件，并让它们覆盖直径 12mm 的孔。将 4 个魔术贴贴在面板底部，以固定底部的线，如图 3-10 所示。

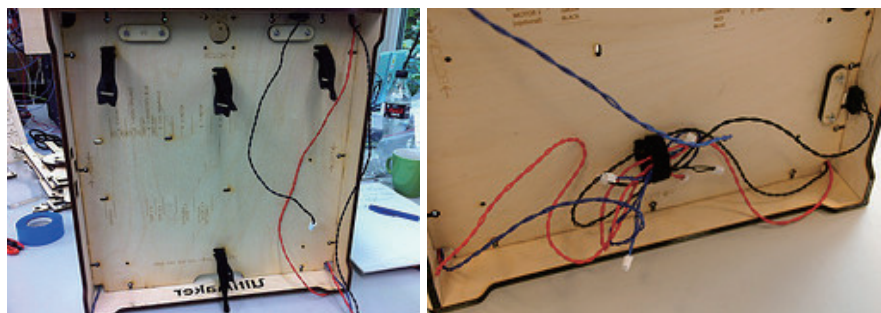


图 3-10 用 4 个魔术贴固定底部的线

3.3.3 X/Y/Z 轴电机

下面，我们安装 X/Y/Z 轴的 3 个电机，如图 3-11 所示。

第 1 步：X 轴和 Y 轴电机装配

此步骤中所需要的零件如图 3-12 所示。



图 3-11 安装 X/Y/Z 电机
(图片来源：Ultimaker wiki)

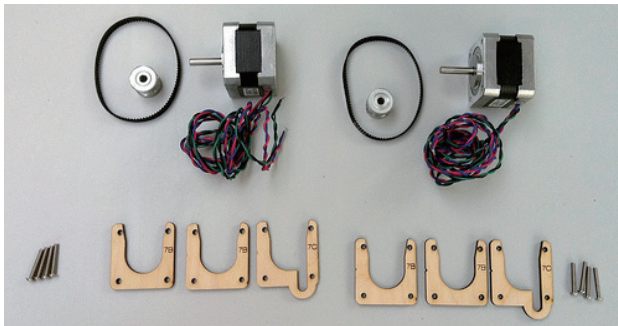


图 3-12 X 轴和 Y 轴电机装配所需零件

1. 准备电机，将内径 5mm 的同步轮放到两电机上。在电机和同步轮之间留有一个微小的间隔 (0.5mm)，并固定好。

2. 下面步骤你需要做两次，一个电机一次。

(1) 将电机背面贴着桌面放置。将零件 7C 放置到电机上面，并将电线理顺，如图 3-13 所示。

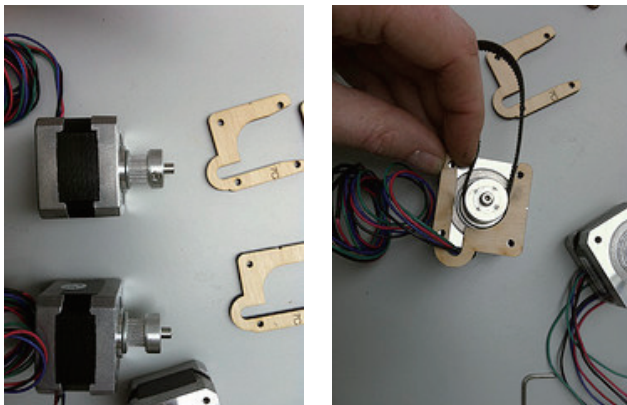


图 3-13 将零件 7C 放置在电机上面

(2) 将零件 7B 放在零件 7C 上面。

第 2 步：安装 X 轴和 Y 轴电机

1. 在电机将要安装的位置，将 4 个带有垫圈的螺栓 (20mm) 穿过机器后背的孔洞，如图 3-14 所示。

2. 放好电机。确保电机的导线、限位开关的导线和同步皮带不被卡住。

3. 稍微拧紧电机的 4 个螺栓,方便往上和往下滑动。等所有轴安装好后,再将这些螺栓拧紧,如图 3-15 所示。Y 轴电机类似,并安装到左面板。在你安装 Y 轴电机前,确保将电缆归拢到角落。

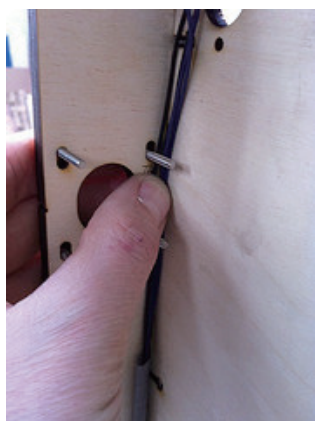


图 3-14 将 4 个带有垫圈的螺栓 (20mm) 穿过机器后背的孔洞

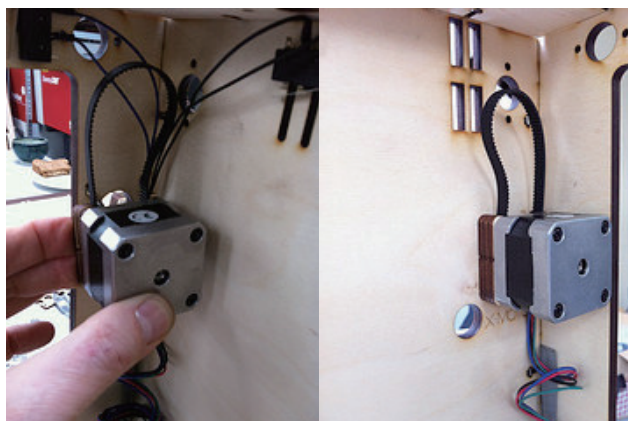


图 3-15 拧紧电机的 4 个螺栓

第 3 步 : Z 轴电机

1. 使用 4 个 10mm 螺栓,将第 3 个电机 (带最短的导线) 安装在底面板的下侧,并确保导线伸出、朝向底面板的中心,如图 3-16 所示。

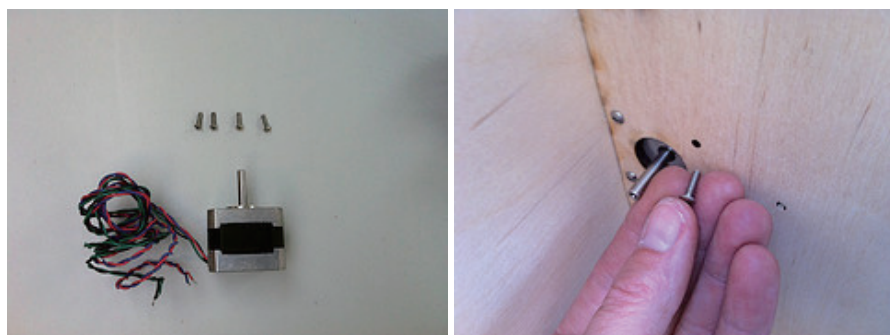


图 3-16 使用螺栓将第 3 个电机安装在底面板的下侧

2. 将联轴器放置到电机顶部。

3. 使用内六角扳手将联轴器设置在正确的高度。联轴器有两个不同的端子,各端子的内径分别为 5mm 和 8mm。

4. 将 5mm 孔径的一端放置在电机顶部并拧紧联轴器底部的小螺钉,如图 3-17 所示。



图 3-17 拧紧联轴器底部的小螺钉

3.3.4 X/Y 轴承

第 1 步：插入球轴承

拿起框架,将 8 个球轴承插入框架顶部的孔中。可用多余的木板将轴承压入,如图 3-18 所示。



图 3-18 将 8 个球轴承插入框架顶部的孔中（图片来源：Ultimaker wiki）

第 2 步：组装 X/Y 套管模块

1. 把轴承插入前面板(FRONT)A 的安装孔中,同样,可使用多余的木板将轴承压入,如图 3-19 所示。

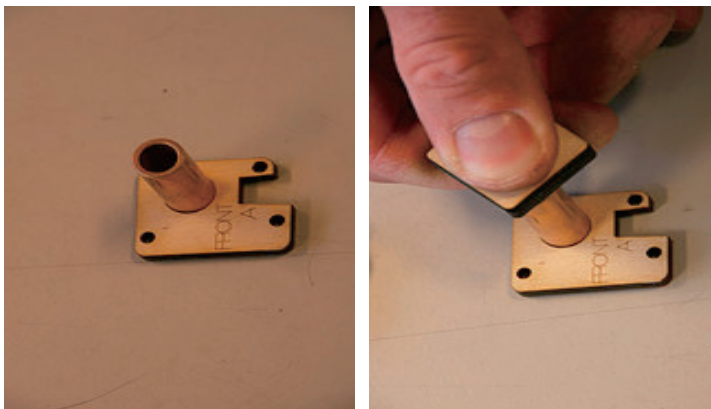


图 3-19 把轴承插入前面板 A 的安装孔中

2. 对前面板 B、前面板 C、前面板 D 和前面板 E 重复相同的操作,直到所有部件都通过球轴承串在一起。

第 3 步：安装卡钳

卡钳是用来卡住同步带的。将 3 个 30mm 的螺栓从有字的一面插入,在另一面拧上螺帽。注意不要拧得太紧,螺丝应该能转动。做好后如图 3-20 所示,下一步就可以组装各个部件了。

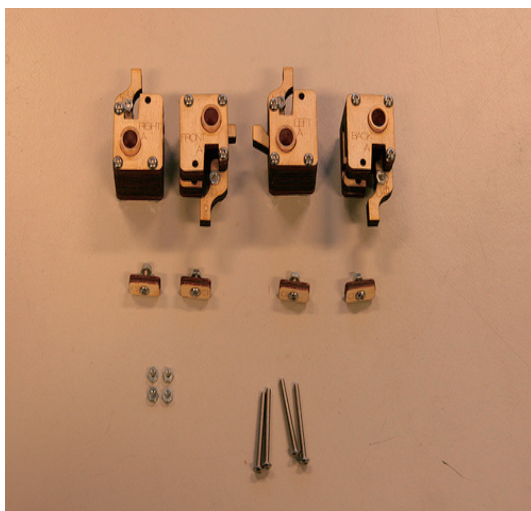


图 3-20 卡钳安装完成后的效果图

第 4 步：组装轴承帽

1. 现在安装轴承帽，这样可以把各个轴固定住，不让它们在操作过程中滑脱。安装轴承帽的顺序很重要。在图 3-21 中，可以看到轴承如何从右侧和后侧插入的。

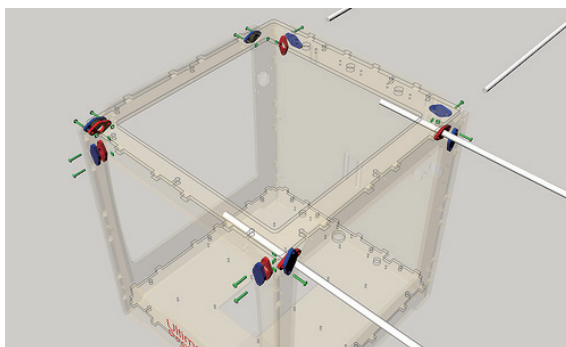


图 3-21 轴承帽的安装

2. 按照下面的顺序安装！从前面板左上方的轴承帽开始安装，外面是一个封闭的轴承帽，里面是一个有孔的轴承帽，如图 3-22 所示。

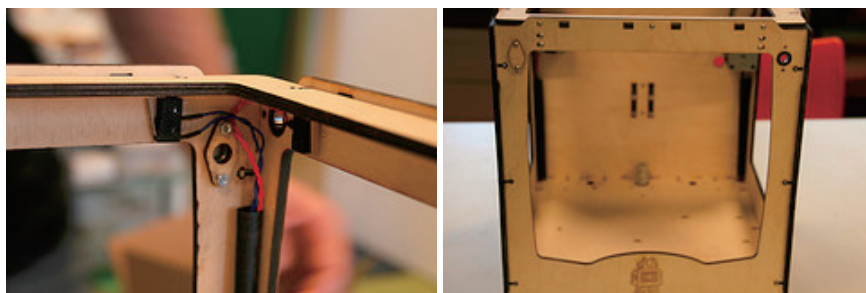


图 3-22 从前面板左上方的轴承帽开始

3. 接下来是前面板右上方的轴承帽。再然后分别是：左面板的轴承帽、右面板的轴承帽、后面板的轴承帽。

第 5 步：安装轴承

1. 检查轴的长度是否合适。有 4 根直径为 8mm 的轴，其中两根短的适合从左到右方向，另外两根稍长的适合从前到后方向。

2. 如图 3-23 所示，取一根同步带绕在同步轮上。然后在轴上串上 FRONT 滑块，带文字的一面正对着机器的 LEFT（左）边。



图 3-23 将轴从前面板插入，然后在轴上串上 FRONT 滑块

3. 然后将另一个同步轮套在轴上，使固定螺丝面向内侧。

4. 另外一根同步带放在同步轮上，然后把轴的另外一端插入面板上的球轴承。旋转轴承帽，并用螺栓固定，用螺帽拧紧。

3.3.5 挤出头

你需要这些部件来安装挤出头，如图 3-24 所示。

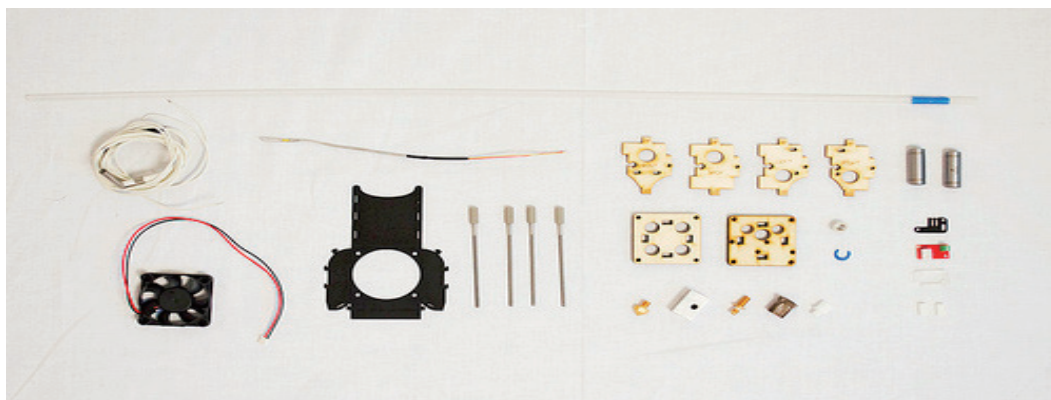


图 3-24 挤出头的部件（图片来源：Ultimaker wiki）

挤出头的最终效果如图 3-25 所示。

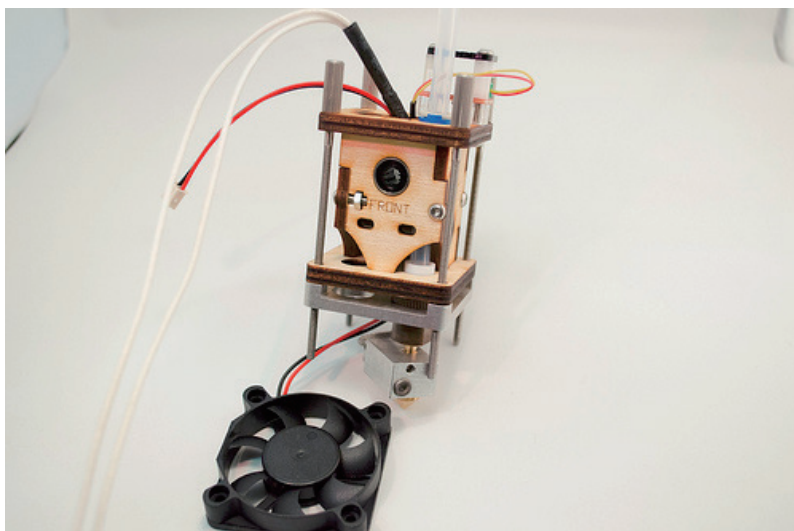


图 3-25 安装好的挤出头

让我们开始吧！

第 1 步：组装挤出头

1. 拿起铝制加热模块,让上面最大的孔位于右下方。拧紧底部的喷嘴和顶部的铜管,如图 3-26 所示。注意铜管插入铝制加热块的部分没有螺纹。

2. 现在拿出标有 FRONT、BACK、LEFT 和 RIGHT 的 4 块木片。拿出 RIGHT 木片放在轴承的一边,另一边是 LEFT 木片。拿起 FRONT 木片放在 LEFT 和 RIGHT 木片之间,如图 3-27 所示。保持刻字的一边朝外。然后拿出 BACK 木片放在 FRONT 木片对面,这样外壳就做好了。用 M3 10mm 螺钉和 T 形槽拧紧外壳。

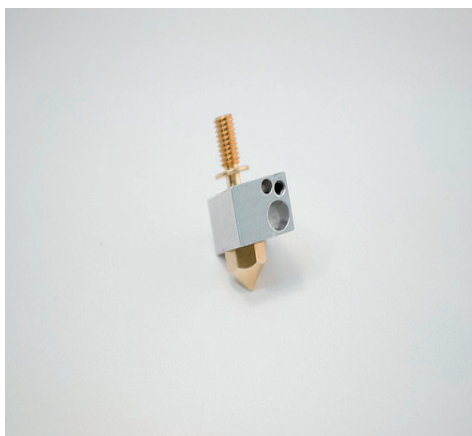


图 3-26 铝制加热模块

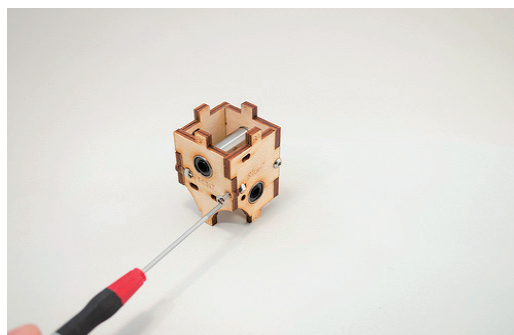


图 3-27 用 M3 10mm 螺钉和 T 形槽拧紧外壳

3. 拿出铝制加热模块(带有刚刚组装上的喷嘴和铜管),将其拧进 PEEK 部件,如图 3-28 所示。确保铜管在 PEEK 部件中不能移动,将其拧紧。

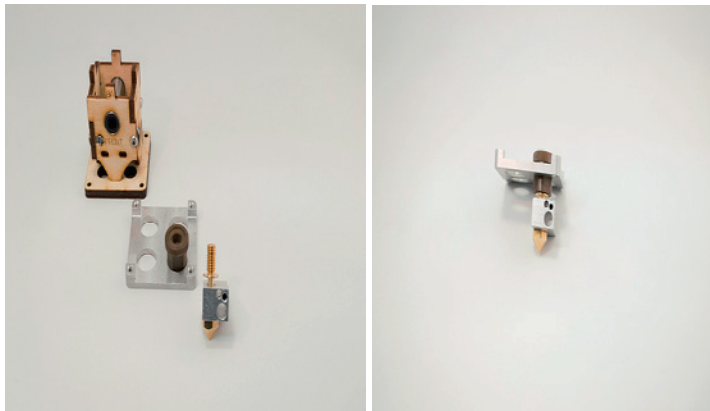


图 3-28 将铝制加热模块拧进 PEEK 部件

4. 将外壳放在旁边。取两个 M3 16mm 螺栓，将它们放在 8A 木质部件下端，穿过亚克力支撑块。确保 3 个开口位于前端，将一端放在桌子上。然后，将放大电路板放在两个 M3 16mm 螺栓上。

5. 拿起一个小螺丝刀，将红色电线拧入 RIGHT 槽，黄色电线拧入 LEFT 槽。如果你要将你的 Ultimaker 升级到双喷头，还需做另外的调整。然后，将彩色的马蹄型卡簧放在下面锁紧进料管。现在将一切都拧紧就 OK 了。

第 2 步：组装冷却风扇

1. 如图 3-29 所示折叠塑料罩。

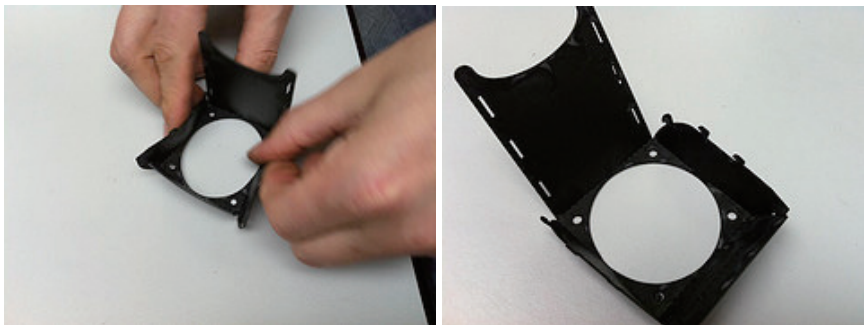


图 3-29 折叠塑料罩

2. 将罩子放置在挤出头左侧伸出的底部，如图 3-30 所示，并用 2 个螺母拧在 M3 螺杆上。

3. 将罩子折叠在一起，同时罩子不应直接接触挤压喷嘴等热元件。你可以通过调整罩子，确保气流不会直接对着喷嘴吹，因为喷嘴不能过分冷却，否则挤压会有问题。

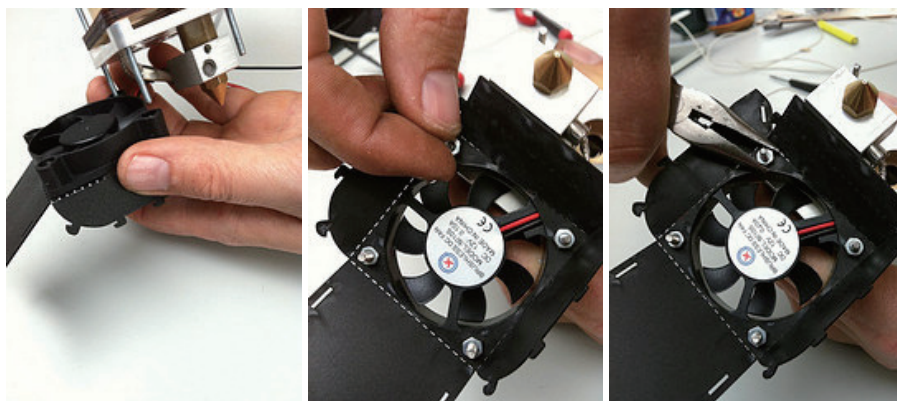


图 3-30 将罩子放置在挤出头左侧伸出的底部

祝贺你！创客大侠！别小看了冷却风扇，你已完成了机器组装中最复杂的一部分，如图 3-31 所示。

第 3 步：将挤出头装到 X-Y 轴上

下面将挤出头装到 X-Y 轴上，本步骤需要准备的部件如图 3-32 所示。

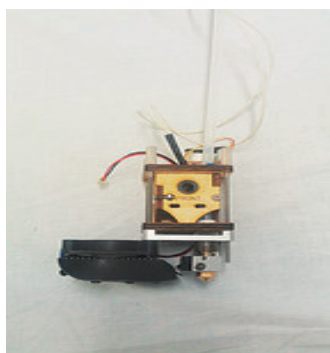


图 3-31 冷却风扇组装完毕

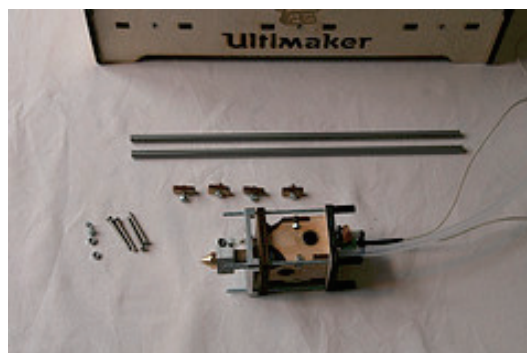


图 3-32 挤出头装到 X-Y 轴所准备好的部件

现在可以将挤出头放置在 X-Y 轴上了。

1. 拿一个带螺丝的零件 C，并把它放到左轴的滑块上，但不要拧紧它，如图 3-33 所示。

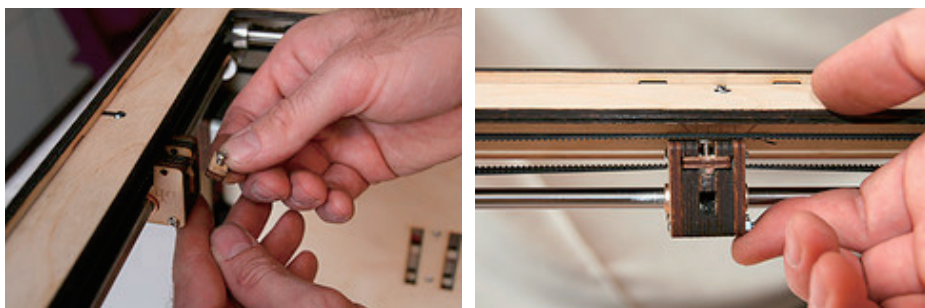


图 3-33 拿一个带螺丝的零件 C，并把它放到左轴的滑块上

2. 取一个 6mm 轴，并将其从左至右地穿过挤出头，确保挤出头的前部朝向机器前部；然后，取另一个 6mm 轴，并将其从前往后地穿过挤出头。保持挤出头在机器里面，并在滑块下面。

3. 最后，调整 X 轴和 Y 轴限位开关。挤出头往一端滑动，并确保你听到限位开关发出“滴答”一声。如果没有滴答声，一点点调整限位开关直到能听到为止。最后拧紧螺栓。所有 4 个都这样做。

3.3.6 Z 轴载物平台

下面我们安装 Z 轴载物平台，安装后的效果如图 3-34 所示。

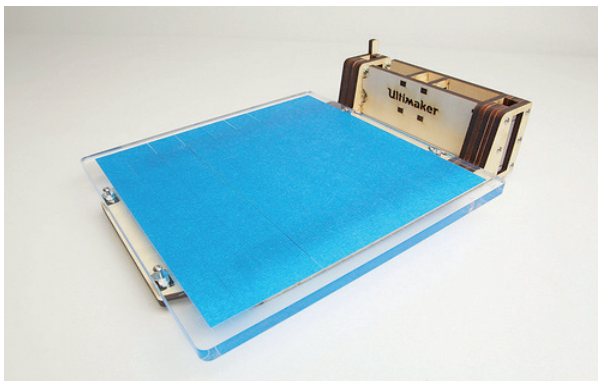


图 3-34 Z 轴载物平台（图片来源：Ultimaker wiki）

第 1 步：驱动螺母的组装

1. 在盒子里找到带轴心的驱动螺母，拧到梯形螺杆上，如图 3-35 所示。

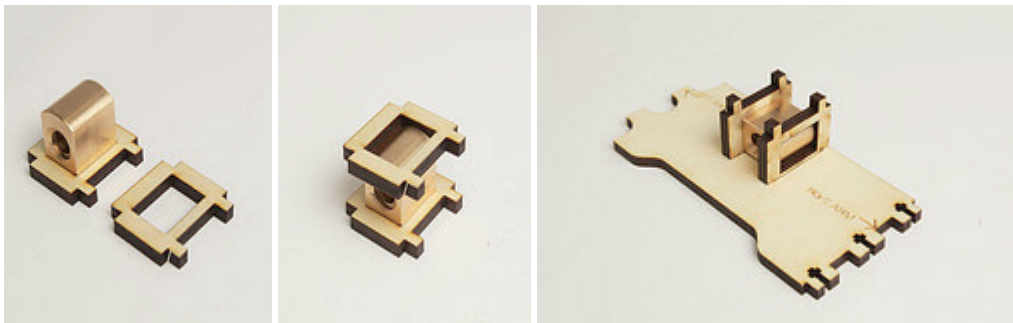


图 3-35 将驱动螺母拧到梯形螺杆上

2. 注意所有文字要朝外。该部件左边有 2 个突起跟左臂配套，右边有 3 个突起跟右臂配套，如图 3-36 所示。

第 2 步：组装左臂

1. 图 3-37 是 Z 轴平台左臂所需的所有部件。拿起 F 部件（让字母朝上）并且将 2 个线轴承中的一个放入。



图 3-36 驱动螺母组装完毕

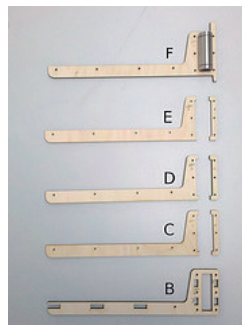


图 3-37 组装 Z 轴平台左臂所需的所有部件

2. 接着将 E 部件放到 F 部件上面，然后是 4D、4C，最后是 4B 部件。

第 3 步：组装右臂

右臂的组装跟左臂类似。组装好的左右臂后的效果应该如图 3-38 所示。

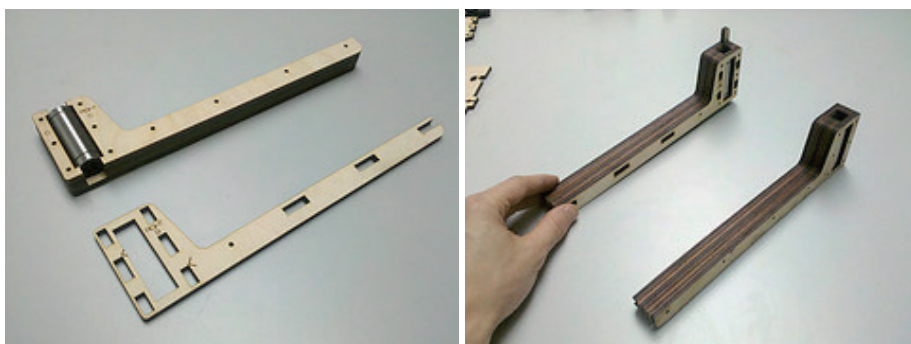


图 3-38 组装好的左右臂

第 4 步：组装所有部件

1. 用 6 个 30mm 的螺栓和六角螺母把左臂固定好。

2. 注意 > 和 < 标记要对好，如图 3-39 所示。



图 3-39 注意对齐 > 和 < 标记

3. 用 4 个 30mm 的螺栓和 4 个六角螺母把右臂接到中央的驱动平板上。3 个六角螺母上到 T 形槽上，1 个六角螺母上到靠近 BACK/CENTER 部件的螺栓上。

4. 在拧紧螺栓前，把平台放在一个平面上。这样当你拧紧螺丝时，平台可以保持水平，然后逐个拧紧边上的螺栓。

第 5 步：插入调整水平的螺钉

用一组弹簧垫圈将整个平台抬高，可通过把螺栓向上或向下拧到黑 / 白树脂部件里来调节。然后，装上亚克力打印平台，如图 3-40 所示，4 个螺栓的顶端穿过平台的孔。

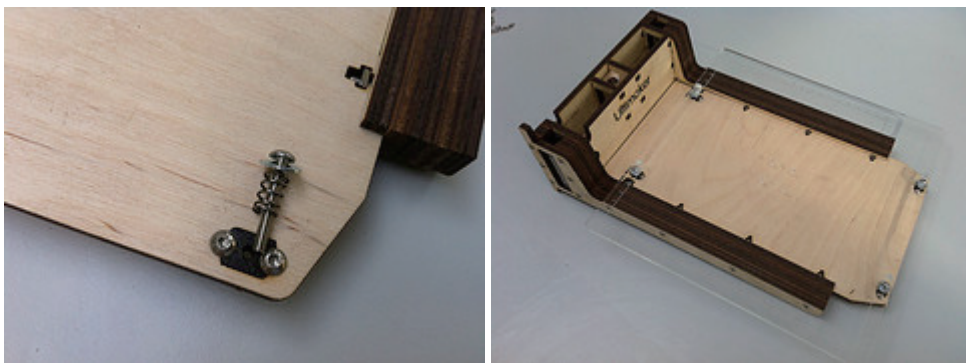


图 3-40 装上亚克力打印平台



提示：你可以用一只手完成大部分工作，用手指抓住左臂的下边缘，大拇指顶住亚克力平台。然后握紧手将玻璃推到位。在右边用右手做同样的操作。

第 6 步：为第一次打印准备好底座

将蓝色胶带贴在亚克力底座上面，从刻在亚克力玻璃上的第一条线开始，然后沿激光刻蚀的凹槽用刀片划一下，去掉多余的胶带，这样就有了一个齐整的打印底座，如图 3-41 所示。

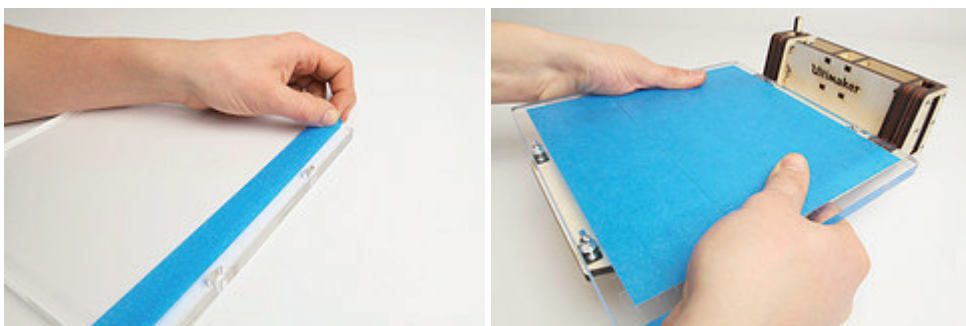


图 3-41 准备好底座

第 7 步：把 Z 轴平台安装在机器上

1. 现在 Z 轴平台可以装在机器里了。将它放在底部，拿一个 12mm 的粗轴线，从机器的顶部插到底，如图 3-42 所示。穿过线轴承，轻轻地推入底部的孔中。

2. 拿起 M8 螺纹,从顶部放入机器。将其拧进 Z 轴平台的螺帽,并一直旋转到 Z 轴平台的底部。现在可以将它固定在底部的联轴器上,用小螺丝刀拧紧。组装好的框架如图 3-43 所示。

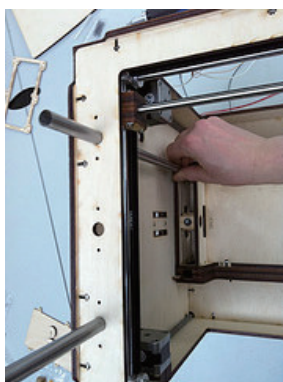


图 3-42 拿一个 12mm 的粗轴线,从机器的顶部插到底

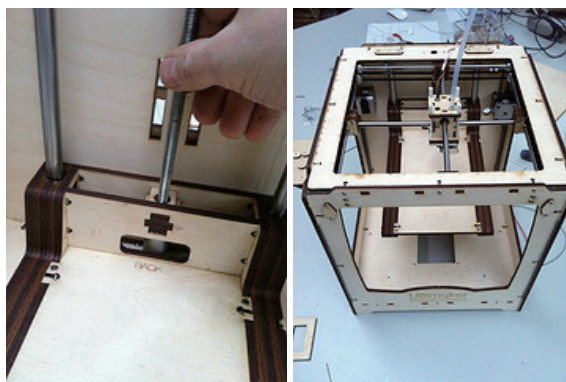


图 3-43 现在拿起 M8 螺纹,从顶部放入机器

第 8 步：涂润滑剂

滴一滴润滑剂在手指上,涂在导向螺丝上,不要涂在其他杆上。然后,当移动 Z 轴时,螺母会将润滑剂涂开。你可以用手转动,或让机器自己转动。用量小于 0.5cm^3 ,大概一颗口香糖大小。

3.3.7 送料机

送料机组装后的效果如图 3-44 所示。

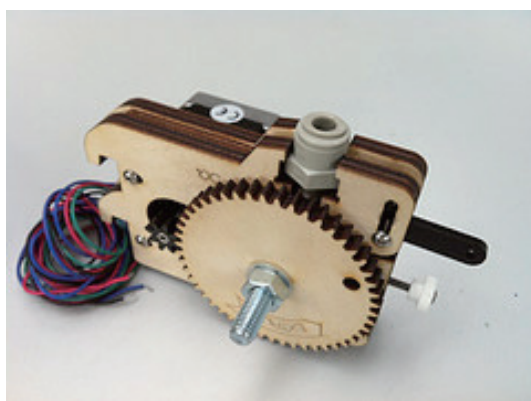


图 3-44 送料机的组装效果图(图片来源: Ultimaker wiki)

第 1 步：安装驱动机的主体

1. 首先准备好木质部件 10A、10B、10C 和电机。
2. 将 10A 放在电机顶部,使文字朝向电机。然后,拿起 10B 放在 10A 上面,将快速耦合器放在 T 形槽内,如图 3-45 所示。

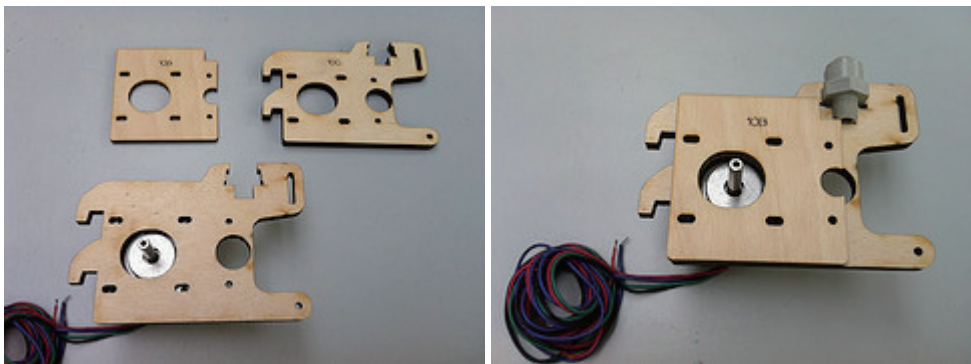


图 3-45 将快速耦合器放在 T 形槽内

第 2 步：驱动螺栓的组装

1. 握住驱动螺栓，把树脂夹子按入凹槽。将其从电机的一端推入轴承，伸到另一端。GOOD 柄部分（接近下面）应该位于两块板材之间。如果不是，检查一下螺栓是否安装牢固，但也不要太紧。

2. 拿起大齿轮，安装上 M8 螺帽。将一个垫圈放在齿轮旁边，再把螺帽松散地拧在螺栓上，然后将齿轮和螺帽拧紧（中间隔着垫圈），如图 3-46 所示。如果齿轮不能转动，而螺栓是静止的，可以在螺帽和齿轮间再加一个垫圈。同时拧紧两个螺帽。



图 3-46 将齿轮和螺帽拧紧

3. 组件的安装顺序以及安装到机器上的效果如图 3-47 所示。

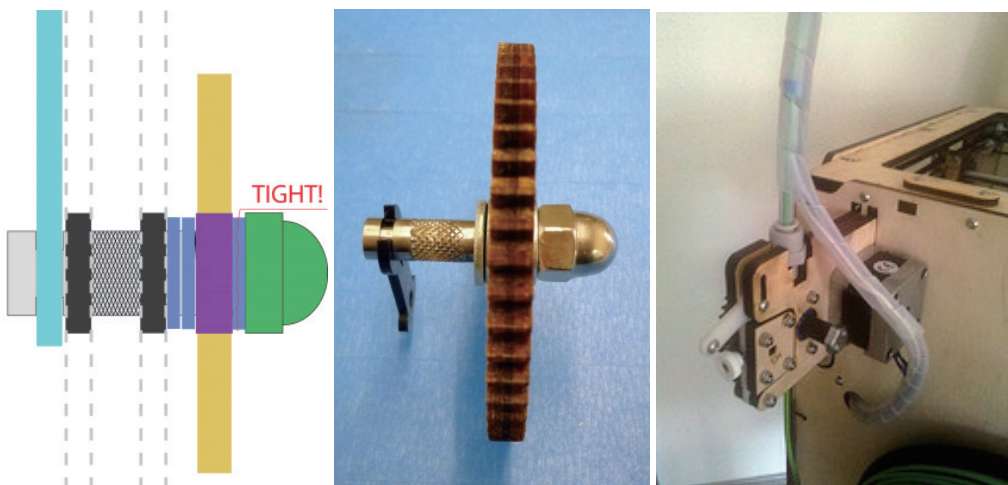


图 3-47 送料机安装完成

3.3.8 Ultimaker 的大脑：电路板

Ultimaker 采用开源的 Arduino 硬件平台，具体采用的主板型号为 Arduino Mega 2560。正因为开源共享，所以才使得 Ultimaker 很快就青出于蓝而胜于蓝了，成为目前桌面级 3D 打印机中的佼佼者。

第 1 步：安装电路板

1. 在底板上从内侧向外侧放置 30mm 螺栓。
2. 将管状垫片放在 30mm 螺栓上
3. 把绿色电路板放在 4 个螺栓上。
4. 用带螺纹的垫片来固定螺栓上的电路板，如图 3-48 所示。

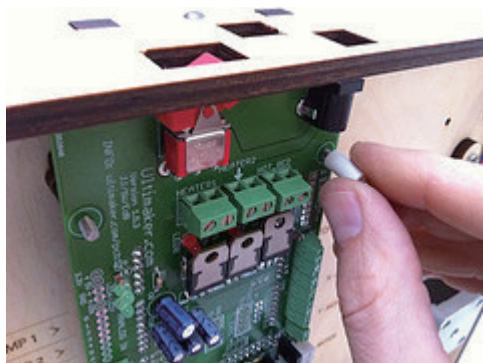


图 3-48 安装电路板

第 2 步：组装电子冷却系统

电路板冷却很重要，过热会导致电路板复位，影响机器的稳定性。安装电子冷却系统所需要的零件如图 3-49 所示。

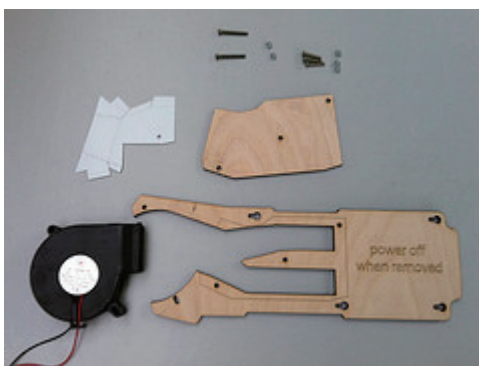


图 3-49 安装电子冷却系统所需要的零件

1. 用 30mm 螺栓，把风扇安装到未标记侧。然后，将冷却传输片放置在该组件的顶部，如图 3-50 所示，并将它沿齿孔折。

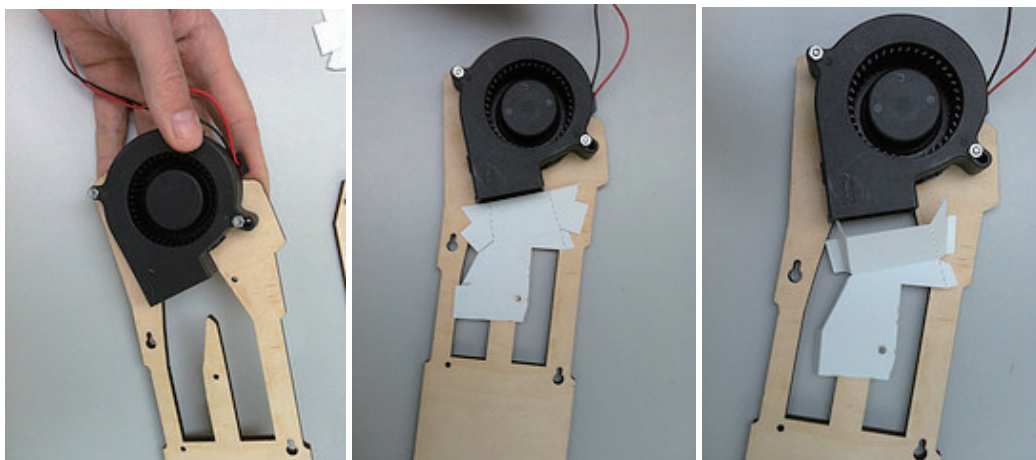


图 3-50 放置冷却传输片

2. 在带标记侧，把木质的空气导板覆盖在开口上面，并用 5 个螺栓（12mm）拧紧。注意要使用短螺栓，这样不会碰到电子器件。然后，将蓝色绝缘胶带贴在螺栓末端以确保它们不会导电。



注意：请一定贴好绝缘胶带，以防止电路板意外损坏。

3. 将风扇连接到 12V 电源输出端（电路板的一个角落上）。插上电源后，当你打开电源开关，风扇应该能够运转。

第 3 步：安装电子冷却系统

1. 将 10mm 螺栓插入白色带螺纹垫片。

2. 把冷却组件放置在螺栓上，如图 3-51 所示，引导它们通过较大的孔。放置到位以便能锁住它。然后，稍微插入螺栓使它们接触木板。

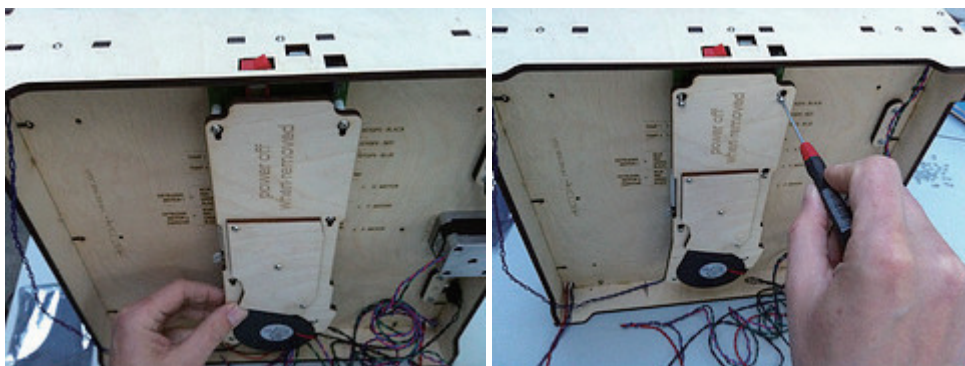


图 3-51 把冷却组件放置在螺栓上

第 4 步：连接加热器

1. 将加热器元件的白色或红色电缆连接到主板上输出加热器的两个端子，如图 3-52 所示。

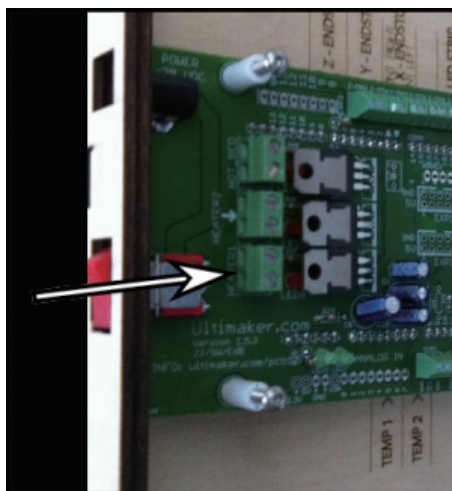


图 3-52 连接加热器元件的电缆到主板

2. 使用小螺丝刀拧紧螺钉接线端子，如图 3-53 所示。

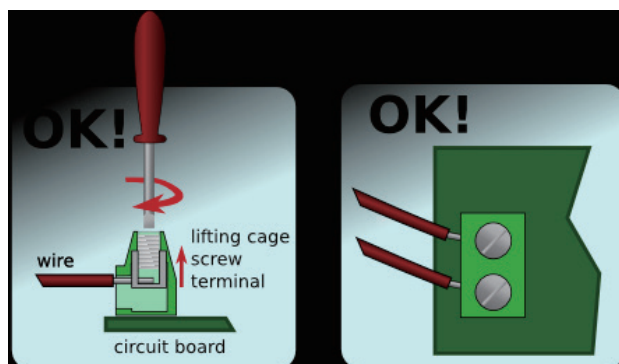


图 3-53 用小螺丝刀拧紧螺钉接线端子

第 5 步：连接打印头的电子器件

1. 使用两侧各带有 3 个插头的电缆（如图 3-54 所示）将电路板和挤出头（带有小电路板）连接起来。



图 3-54 用电缆将电路板和挤出机机头连接起来

2. 然后将限位开关和电路板连接起来。每个限位开关都有一个号码，给出了下一个限位开关和限位开关的头所在的位置。

第 6 步：连接电机

现在可以将电机和电路板用线连接在一起了，如图 3-55 所示。

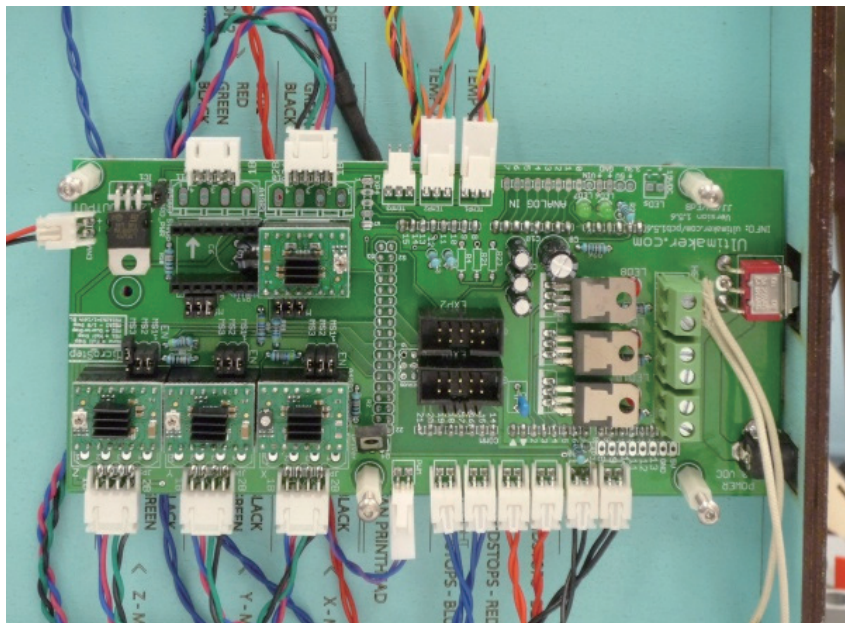


图 3-55 将电机和电路板用线连接在一起

3.3.9 大功告成：一台完整的打印机

把杂乱的电缆用缠丝或扣带整理成线束，以显得整齐划一。这样，一台完整的 3D 打印机就组装好了！这里，以国内一款标牌为 DreamMaker 的打印机（国内类似的还有 Ultimaker 3DJoy、SikMaker 等机型）为例，展示了 Ultimaker 打印机组装完成后的样子，如图 3-56 所示。

目前市面上已有上百种个人 3D 打印机型号，给人一种眼花缭乱的感觉，这里对它们小结一下。

首先，框架结构分为**矩形盒式结构**（MakerBot、Ultimaker、Mbot）、**矩形杆式结构**（PrintrBot）、**三角形结构**（RepRap）、**三角爪式结构**（Rostock）、**舵机转动型结构**。它们所对应的 X/Y/Z 机械轴空间定位方式，在数学上可归结为**笛卡儿坐标系**（直角和斜角）和**极坐标系**两大类。

其次，在控制电路板部分，要么采用**Arduino Mega 开源集成套件**（主要代表有 RAMPS, RepRap Arduino Mega Pololu Shield, **Ultimaker Electronics**），要么**直接用 AMTEL ATMEGA644P、ATMEGA1284 等芯片**将单片机和控制电路做在一起（主要代表有 Sanguinololu、Printrboard、Melzi、GEN 6、GEN 7）。前者很方便扩展，但因为集成了很多资源接口，所以初次投入成本稍高，体积也稍大一些。后者体积较小、初始成本稍低，但因为不是集成套件，所以后期维护比较困难。

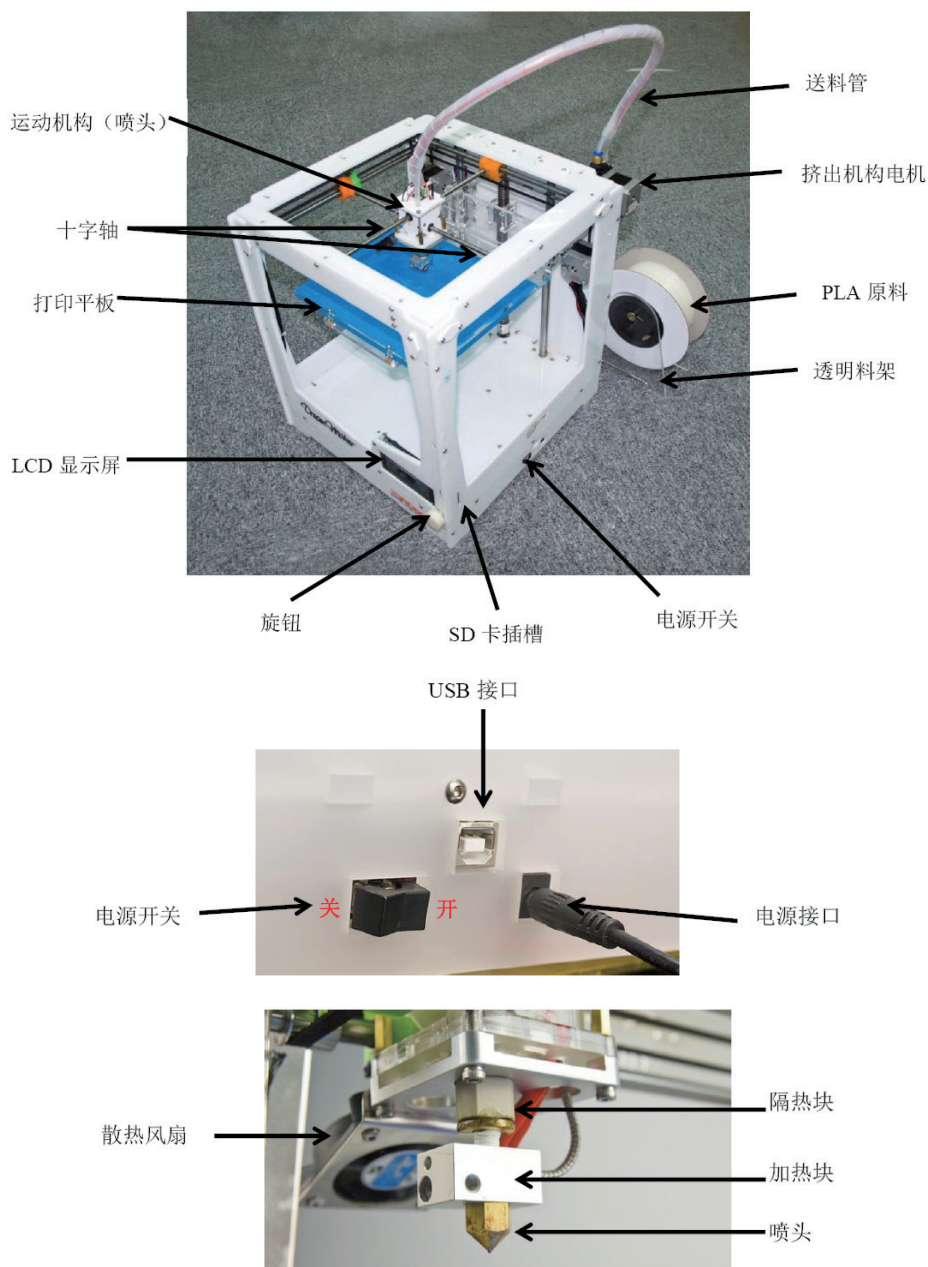


图 3-56 Ultimaker 打印机组装完成后的样子 (图片来源 : DreamMaker)

喷嘴主要分为两种：J-head 和 Budaschnozzle。J-head 重量较轻，适合用在精度要求较高，或者机械轴负载能力较弱的结构中，比如三角爪式结构。Budaschnozzle 喷头有主动散热和被动散热两种方式，比如 MakerBot、RepRapPro 上的 MK7 喷头便采用主动式散热。

挤出机主要分为直接挤丝（Direct Driver Extruder）、齿轮挤丝（Wade's Extruder）和液体挤出 3 种类型。

总之，上面这些术语对于 DIY 技术爱好者有帮助，普通用户可直接忽略之。

3.3.10 Gcode 与前台软件 Cura 使用指南

机器的机械装置和硬件电路已经组装完毕，下面我们正式进入软件的安装和使用教程。

这里我们介绍一款名为 Cura 的**前台控制软件**，也被称为**上位机控制软件、主机软件**，是**运行在计算机**而不是打印机上的软件，其他更多的前台控制软件还有 MakerWare、ReplicatorG、Repetier-Host、Printrun（含 pronterface）、Repetier-Server、BotQueue 等。

Cura 包含了**切片软件工具**（比如 CuraEngine、Slic3r、Skeinforge、KISSlicer、SFACT），可对 3D 模型文件进行**分层切片（Slice）**，以导出可被 3D 打印机识别的**Gcode 控制文件**。然后，我们将文件复制到 SD 卡，插入到 3D 打印机的卡槽中。



提示：STL 模型（图 3-57（a））的**分层切片**（图 3-57（b））处理就是根据分层方向和分层厚度，**求取一系列切平面与 STL 模型中三角面片的交线**，并将首尾相连的线段**组成截面轮廓**，同时还要判断轮廓是否**封闭**。分层处理一般包含两个步骤：即**平面求交**和**线段轮廓归并**。具体地，在求取每一层的轮廓线段时，需判断**每个三角形面片与切平面的位置关系，若相交则求交线**。在完成求交运算后，对所得交线进行**排序**，以生成**封闭轮廓线**（图 3-57（c）中的轮廓边线）。切片完成之后，还需进一步对**每层截面进行扫描**，方式有**顺序往返直线扫描**、分区扫描、环形扫描、分形扫描、三角剖分扫描等，**以便对一个截面轮廓的内部进行扫描填充**（图 3-57（c）中内部的密集横线）。最后，所有层被叠放在一起，生成完整的**Gcode**，就可以打印出一个立体的模型（图 3-57（d）为打印机挤出头行走的路径图）。

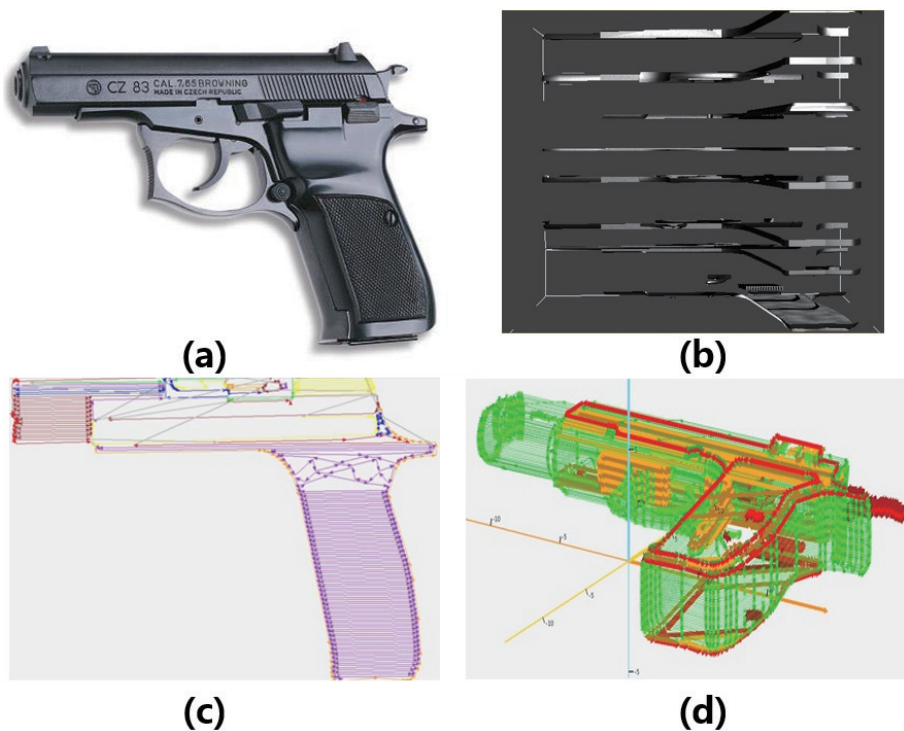


图 3-57 对一把手枪模型进行分层切片的原理图（图片来源：demohour.com/1045705）

最后，固化在3D打印机里Arduino主板上的软件（即**固件**，目前有很多软件版本，如Marlin、Sprinter、Sailfish）便开始读取SD卡上的Gcode文件，并以此控制电机开始逐层打印3D模型了。

以上整个流程，形成了“**固件→前台控制软件→切片软件工具**”这样一个工具链（**Tool Chain**），你可以对每个环节的工具进行自由选择 and 搭配，比如“Marlin→pronterface→Slic3r”就是一种常被选用的工具链。



提示：Gcode文件包含了控制打印机动作的完整指令步骤，以完成某个3D模型（如一只鸡）的打印。形象地说，Gcode就像一个菜谱，比如我们要制作一道菜“宫保鸡丁”，它把什么时候需放什么料（以及进料速度）、某个阶段是大火还是温火（精确到多少摄氏度），每个翻炒动作（精准到手拿铲子的姿势角度）以及每个步骤的持续时间（精确到ms）等都详细地写在里面，以确保任何人拿着这本菜谱炒出的菜都是完全一模一样的口味！

比如，一句能够被Ultimaker打印机识别的Gcode代码如下（；号后面为解释）：

```
N3 T0*57 ; 前面的N3为行号，后面的*57为校验码，T0代表选择第0个喷头
N4 G92 E0*67 ;G92设置位置，E0设置喷头的Z坐标位置为0
N5 G28*22 ;G28代表移动到原点
N6 G1 F1500.0*82 ;G1代表喷头移动前，先设置每分钟1500mm的进料速度
N7 G1 X2.0 Y2.0 F3000.0*85 ; 从当前XY坐标移动到目的点(2.0, 2.0)，同时改变进料速度为每分钟3000mm
N8 G1 X3.0 Y3.0*33 ; 从当前XY坐标移动到目的点(3.0, 3.0)
```

怎么样？Gcode是不是也很简单？在实际使用过程中更简单，因为软件（如Cura）会自动为STL模型生成相应的Gcode，**普通用户根本就不需要阅读Gcode代码**！

导出Gcode文件后，将它复制到SD存储卡再输入打印机，3D打印机就可以开始制作实体模型了。下面我们重点介绍如何用Cura软件生成Gcode文件，以及各种高级设置。

Cura 的安装

双击Cura安装程序进行安装，选择安装位置。注意：目录名及文件名均为英文字符，因为该软件暂不支持中文字符；然后，选择需要安装的组件，如图3-58所示，如果你希望还可使用OBJ格式文件，请选中“Open OBJ files with Cura”复选框。

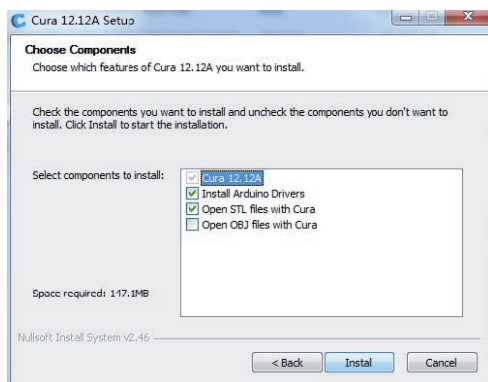


图 3-58 Cura 安装界面

单击“Install”按钮开始安装。如果你是第一次安装该软件，软件会提示安装 3D 打印机的驱动程序。最后单击“完成”按钮。恭喜，你已顺利地安装好了 Cura。

Cura 的设置

Cura 的初始界面如图 3-59 所示，在这个界面中你可以简单地选择“打印质量（Select a print type）”、“打印耗材（Material）”和“是否打印支撑（Other: Print support structure）”，在右侧界面中我们可以预览打印后的模型。

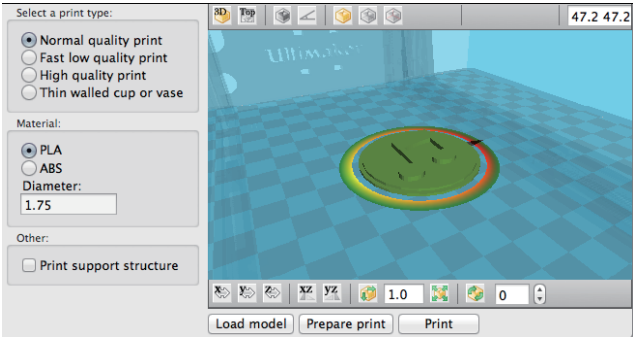


图 3-59 Cura 的初始界面

如果需要对打印效果做进一步的修改，可以在“Tools”菜单下选择“Switch to full settings...”对各项高级参数进行修改。以下就是完全设置模式，按照图 3-60 中的值进行设定，一般都能获得较好的打印效果。

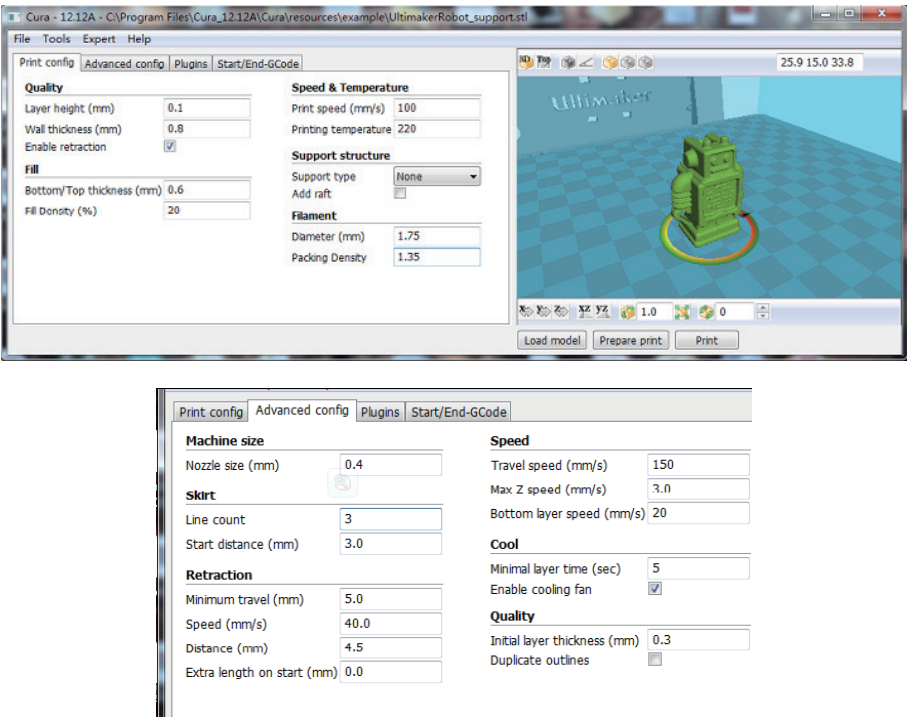


图 3-60 完全设置模式的各种参数

Cura 的使用——得到可打印的 Gcode 文件

设置完成 Cura 后，即可使用 Cura 导出 STL 文件的 Gcode 代码。

过程很简单，单击“Load model”按钮，选择你需要打印的 STL 文件，如图 3-61 所示。

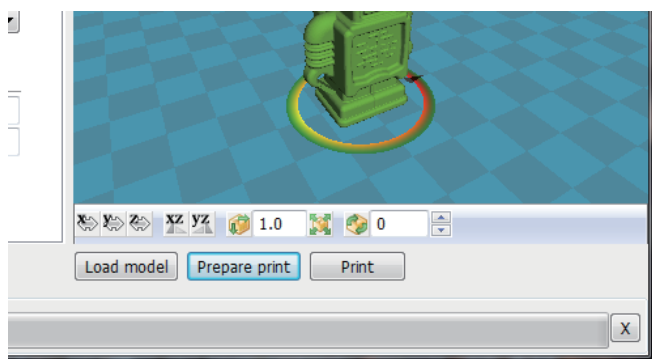


图 3-61 单击“Load model”按钮选择要打印的 STL 文件

再单击“Prepare print”按钮，软件自动在 STL 文件的同一目录下生成同名的 gcode 后缀文件，如图 3-62 所示。

名称	修改日期	类型	大小
Attribution.txt	2012/12/24 17:13	TXT 文件	1 KB
UltimakerHandle.stl	2012/12/24 17:13	STL 文件	439 KB
UltimakerRobot_support.gcode	2013/3/1 14:18	GCODE 文件	2,141 KB
UltimakerRobot_support.stl	2012/12/24 17:13	STL 文件	9,179 KB

图 3-62 生成与 STL 文件同名的 gcode 后缀文件

将 Gcode 文件打印成 3D 实体模型

将 Gcode 文件复制至 SD 存储卡，即可放入 3D 打印机进行打印了。

首先给打印机接上电源，如图 3-63 所示。连接打印机和标配电源适配器，将电源适配器插在 100~240V 插座上。

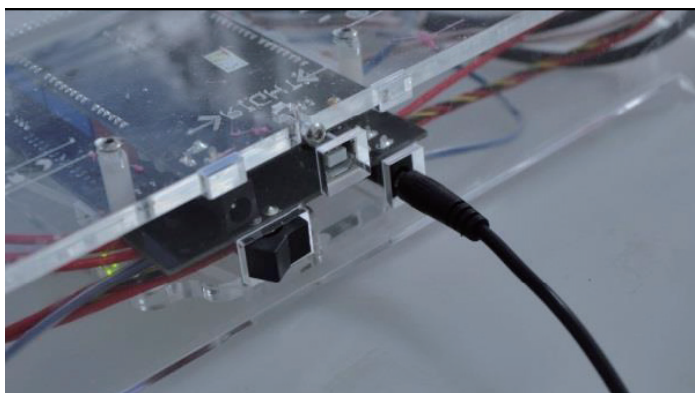


图 3-63 给打印机接上电源（图片来源：DreamMaker）

打开打印机右侧外壳上的电源开关，在显示屏右方卡槽内插入 SD 卡，如图 3-64 所示。

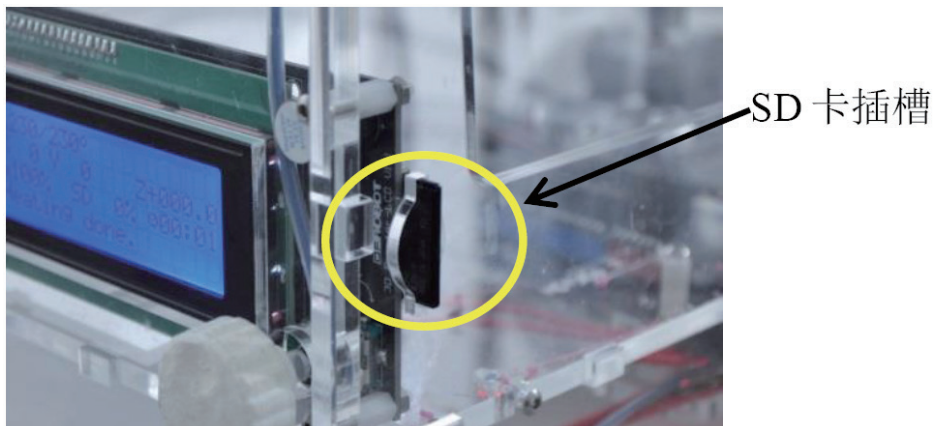


图 3-64 在显示屏右方卡槽内插入 SD 卡

打印机上电后的 LCD 屏如图 3-65 所示。为加快准备时间，可提前预热。在待机状态下按压控制面板旋钮，进入菜单。顺时针转动旋钮，选择“Prepare”菜单，并按压旋钮；接着顺时针转动旋钮，选择“Preheat PLA”，并按压旋钮。此时 LCD 上将显示设定的温度值和当前值。一般 PLA 材料的预热温度为 220℃。



图 3-65 选择“Prepare”菜单进行预热

在待机状态下，按压旋钮进入菜单。顺时针转动旋钮，选择“Card Menu”，按压旋钮进入 SD 卡菜单。顺时针转动旋钮，选择已经生成的 Gcode 文件，按压旋钮确认打印，如图 3-66 所示。

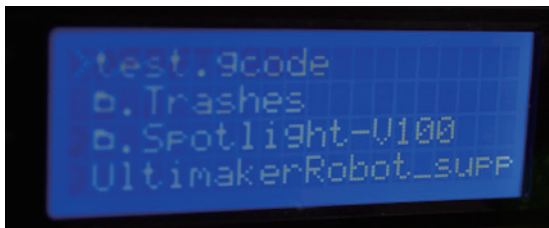


图 3-66 选择 Gcode 文件确认打印

进入打印状态后，打印机会自动进行加温。请等待加温，待温度达到后将会进入打印状态。此时 LCD 显示屏会显示打印进度、打印时间、打印速度等参数。

恭喜你，已经让 3D 打印机开始打印了！

3.3.11 Ultimaker 打印成果实例

当打印过程结束后，喷头会自动归位。但比较头疼的是 3D 模型会“牢牢粘在”打印平板上。可参照下面的方法使用小铲子将模型与打印平板剥离。首先，将打印平板向螺丝孔缺口方向平移，向上摘除打印平板。用小铲子小心插入打印好的模型底部，并轻轻地往上翘起。当翘起一两毫米之后，需要寻找另外一个对角或者附近未被翘起的底部继续将铲子插入，再次一点点地翘起，直到将整个模型剥离，如图 3-67 所示。

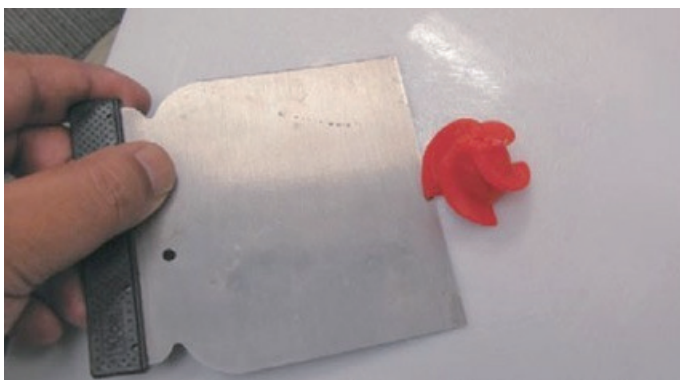


图 3-67 用小铲子将模型与打印平板剥离（图片来源：中日技术产业信息网）

Ultimaker 的各种打印案例如图 3-68 所示，是不是细节表现得很精致？

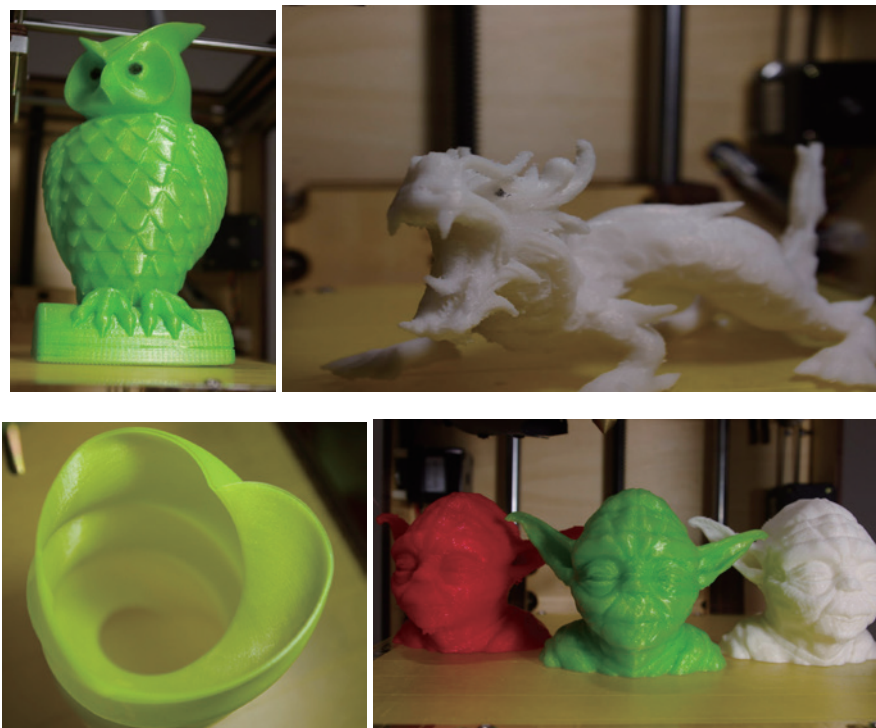


图 3-68 Ultimaker 的打印案例，使用 PLA 材料（图片来源：JoysMaker）

3.4 MakerBot Replicator 2与MakerWare打印实战

在 3.3 节我们已经介绍了 Ultimaker 的组装和打印实战。而作为“桌面双雄”的另一雄——MakerBot，也是目前市面上响当当的主流机型。因此我们也对 MakerBot 的打印操作做一个介绍。

将 MakerBot Replicator 2 接通电源。初次使用时，打印机的液晶显示屏会提醒用户按照向导进行底板的调平。可用肉眼观察打印头与底板之间的距离，如需调整，底板下面有 3 个可手动旋转的螺丝，如图 3-69 所示，通过旋转来调至底板水平。具体来讲，用手旋转 3 个调平螺丝，使打印头喷嘴与底板的空隙均匀，空隙以放入一张纸即为合适。

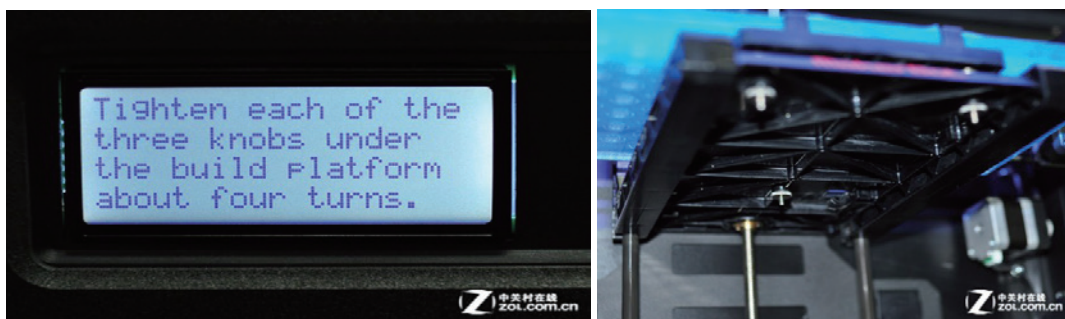


图 3-69 MakerBot Replicator 2 底板调平向导、底板下的 3 个调平螺丝（图片来源：ZOL）

3.4.1 MakerWare 进行切片和打印

类似于 Ultimaker 平台下的 Cura 软件，MakerBot 也有一款前台软件：MakerWare。MakerWare 包括新的切片工具 Miracle Grue（现已改名 MakerBot Slicer），速度可以达到 ReplicatorG 中切片工具 Skeinforge 的 20 倍。MakerWare 软件主要做的事就是让打印机知道该打印什么东西，打印参数的修改，还有模型大小的缩放等。与众不同的是，这款软件能够同时打开多个 3D 模型文件进行设置，可以同时打印多个模型，当然也可以单个分别进行打印。除此之外，MakerWare 还有一个功能，就是当用户调整完要打印的模型之后，可以将参数导出为一个特定的 thing 文件，以便用户下一次直接使用。

MakerWare 的界面十分简洁，软件打开速度也很快，主界面上的 12 个硕大的按键让我们不需要花很多时间就可以了解它们的功能。下面是具体的使用方法。

我们将用一个例子教你使用 MakerWare，例如，同时打印两个名为“Flatiron Building”（熨斗大楼，纽约第一座摩天大楼，因形似熨斗而得名）和“Woolworth”（伍尔沃斯大楼，纽约市的另一座摩天大楼）的 3D 模型。在 Thingiverse 网站查找到“FlatIron.stl”和“Woolworth.stl”，并将它们下载到你的计算机。

打开 MakerWare。单击窗口顶部的“Add（添加）”按钮，在弹出的对话框中选取“FlatIron.stl”文件。3D 模型将出现在中心的网格灰色区域，如图 3-70 所示，该区域代表着实际的打印平台，长方形线框指示了可打印区域。

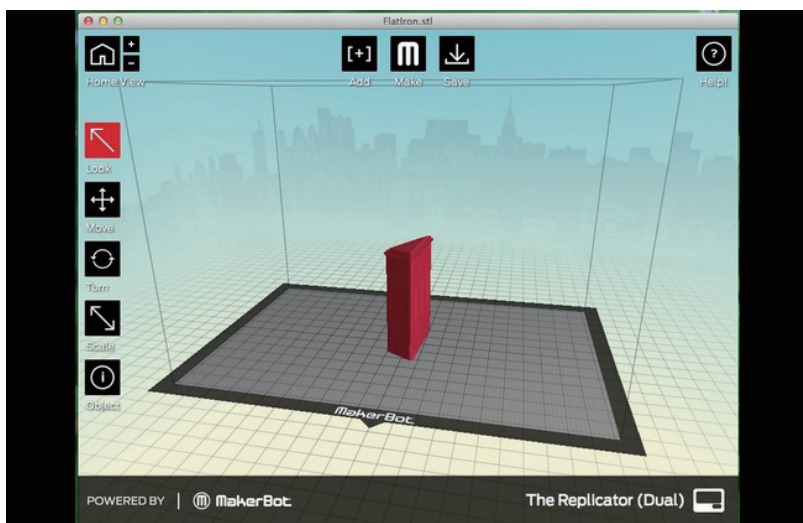


图 3-70 MakerWare 的主界面

现在，你有了一个打开的 3D 模型，我们来看看有哪些选项可供查看和操作。

- Home View。将场景恢复到默认视图。
- +/-。放大和缩小。你也可以使用鼠标上的滚轮进行放大和缩小。
- Look (查看)。单击“Look”按钮或按 L 键进入查看模式。在这种模式下，拖动鼠标来旋转面板和物体。单击出现在“Look”按钮右侧的小箭头 ►, 打开“Change View(更改视图)”子菜单，如图 3-71 所示，以查看顶部、侧面或正面视角。

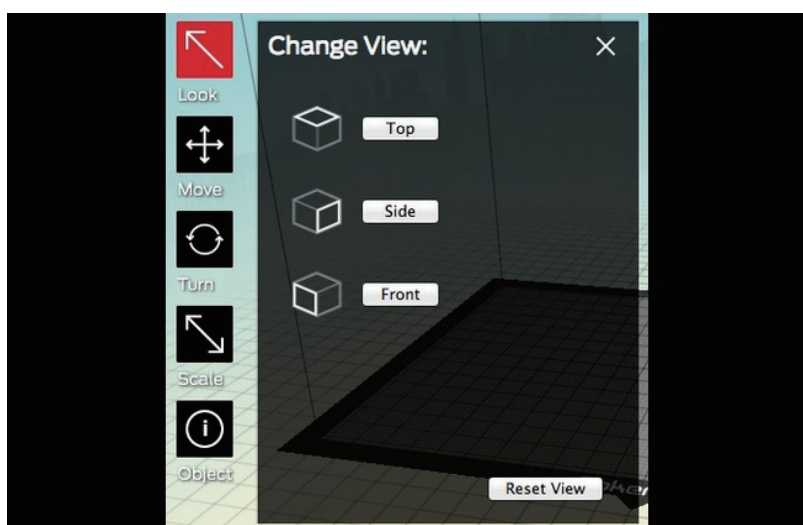


图 3-71 Change View (更改视图)

- Move (移动)。单击“Move”按钮或按 M 键进入移动模式。在这种模式下，拖动鼠标来移动一个物体。单击出现在“Move”按钮右侧的小箭头, 打开“Change Position(更改位置)”子菜单来设定对物件的移动距离，如图 3-72 所示。

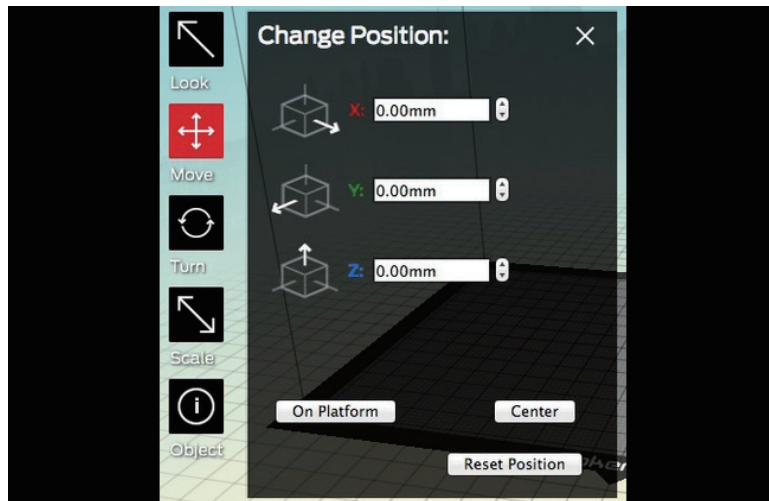


图 3-72 Change Position (更改位置)

- Turn(旋转)。单击“Turn”按钮或按 T 键进入旋转模式。在这种模式下,拖动鼠标来旋转物体。单击出现在“Turn”按钮右侧的小箭头,打开“Change Rotation (更改转角)”子菜单来指定物件的旋转角度,如图 3-73 所示。
- Scale (缩放)。单击“Scale”按钮或按 S 键进入缩放模式。在这种模式下,拖动鼠标来放大或缩小物件。单击出现在“Scale”按钮右侧的小箭头,打开“Change Dimensions (更改尺寸)”子菜单,如图 3-74 所示。
- Object(对象)。单击此按钮打开“Object Information(对象信息)”子菜单。在这个菜单中,你可以指定每个物件对象由哪个挤出头(喷头)打印。该按钮只在机器配有双挤出头时才会出现。

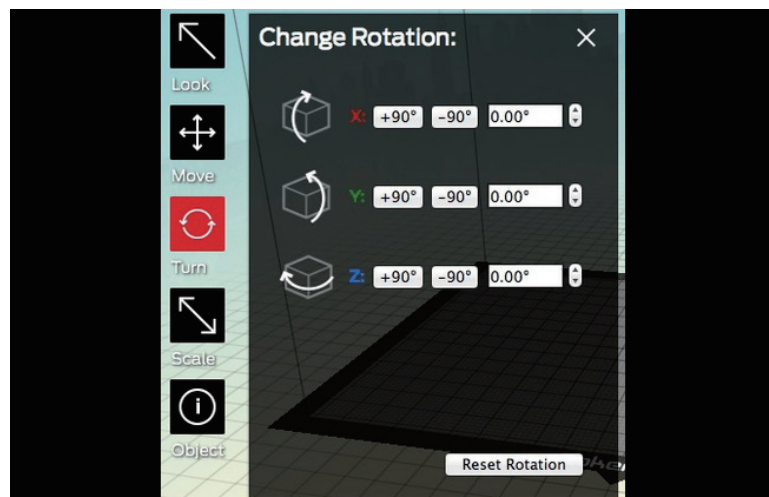


图 3-73 Change Rotation (更改转角)

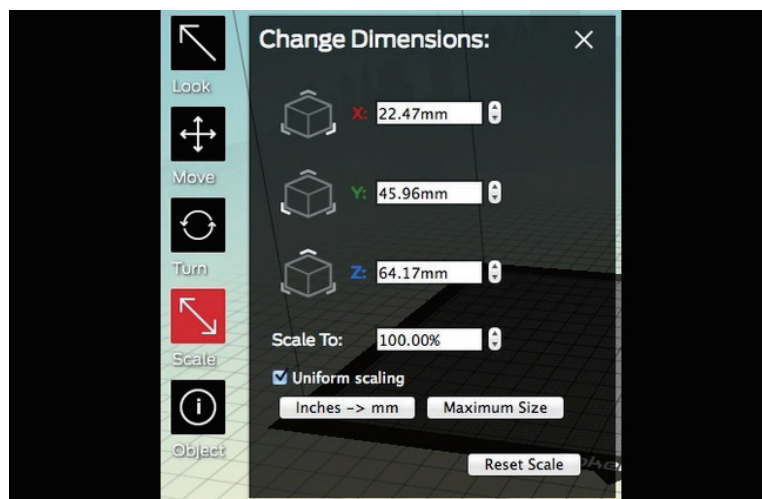


图 3-74 Change Dimensions (更改尺寸)

- Add (添加)。单击此按钮将另一个模型添加到你的打印平台。你可以添加尽可能多的物件，只要平台上放得下。
- Make it (开始打印)。单击此按钮可打开“Make”打印对话框，在这里你可以指定打印分辨率并将你的物件发送到 MakerBot 进行打印。
- Save (保存)。保存当前打印平台的参数为文件，供以后使用。
- Help (帮助)。帮助向导。
- Status bar (状态栏)。显示与 MakerBot 打印机的连接状态和打印进程。

下面，我们要添加第 2 个 3D 模型，所以我们首先需要把第一个物件移开。选择“移动”按钮，然后把 Flatiron 大厦模型从中心拖动到左边，如图 3-75 所示。

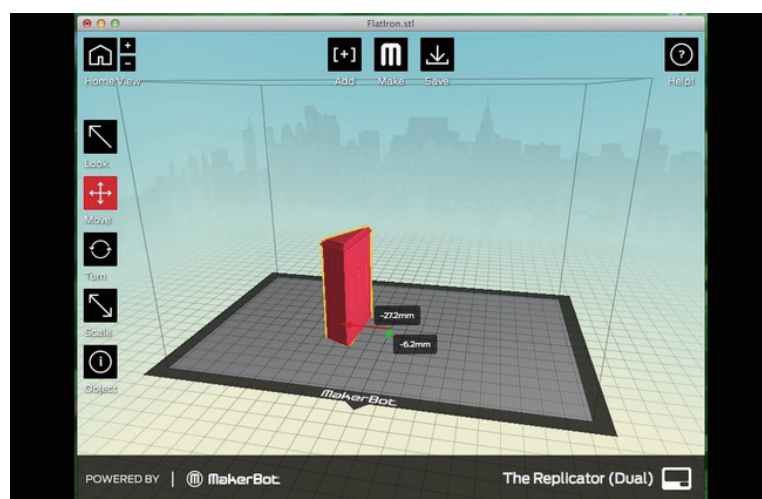


图 3-75 将 Flatiron 大厦从中心拖动到左边

单击“Add (添加)”按钮，在弹出的对话框中选择“Woolworth.stl”文件。现在，你可以看到 Flatiron 大厦和 Woolworth 大厦模型都显示在虚拟的打印平台上。



提示：你也可以复制已有的物件。只要选择好对象，先按 Ctrl + C 组合键进行复制，再按 Ctrl + V 组合键进行粘贴，就得到了两个一模一样的模型。

打开多个模型后，你可以单独或一起操纵它们。比如，选择其中一个模型，然后单击“Turn（旋转）”按钮或使用 T 键来旋转它，如图 3-76 所示。

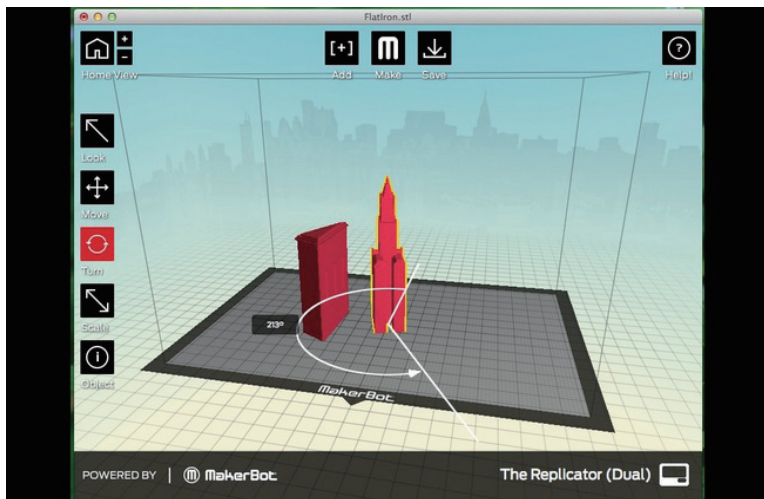


图 3-76 单独旋转一个模型

或者，我们可对两个模型一起操作。单击 Flatiron 大厦模型以选择它，然后按住 Shift 键并用鼠标单击 Woolworth 大厦模型，最后松开 Shift 键，这样两个模型会被同时选中。单击“Scale（缩放）”按钮，拖动鼠标来同时改变这两个模型的大小，如图 3-77 所示。

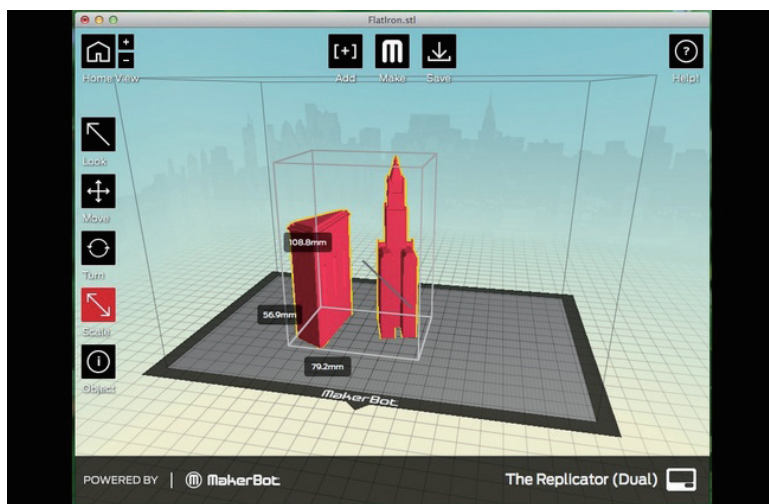


图 3-77 同时缩放两个模型

单击“保存”按钮，在弹出的对话框中保存为 STL 或 thing 文件格式。例如，你可以将文件命名为 Flatiron_Woolworth.thing，如图 3-78 所示。

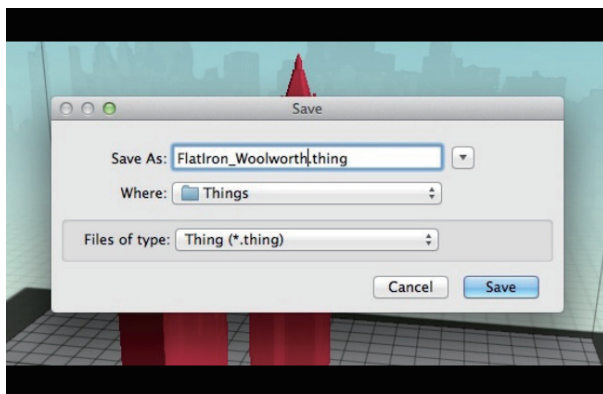


图 3-78 保存场景文件

如果你希望现在就开始打印，可单击“Make it”按钮，弹出如图 3-79 所示的对话框。

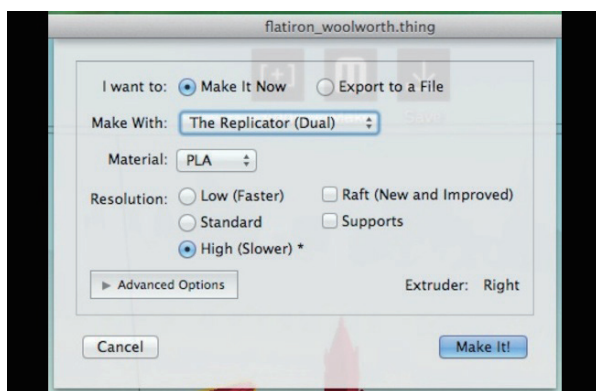


图 3-79 打印选项

Make with。在下拉菜单中指定打印机名称。如果你的 MakerBot 已连接到计算机，它应该被自动选中。



提示：如果你希望先复制到 SD 存储卡再插入打印机进行打印，而不是直接从 MakerWare 打印，则应选择“Export to a File（导出到文件）”，而不是“Make it Now（现在就开始）”。SD 卡打印是大多数情况的选择，因为打印往往要耗费几个小时到几十个小时，而且打印时会产生噪声和散发一些味道，所以目前的 3D 打印机一般都会放在单独的工作间，而不与计算机直接相连。

- Material。选择打印耗材的类型。
- Quality。指定打印的质量。层厚越小，则生成的模型越精细，但更耗时。
- Raft。选中此复选框，则将对象打印在一个 Raft（筏、打印底座）上。如果打印平台不是水平的话，Raft 可以提供辅助支撑。打印完成后，你可以很轻松地将其剥离。
- Support。如果你的物件形状有悬垂部件，选中此复选框，则打印额外的辅助支撑结构。同样，打印完成后，你可以很轻松地将其剥离，如图 3-80 所示。

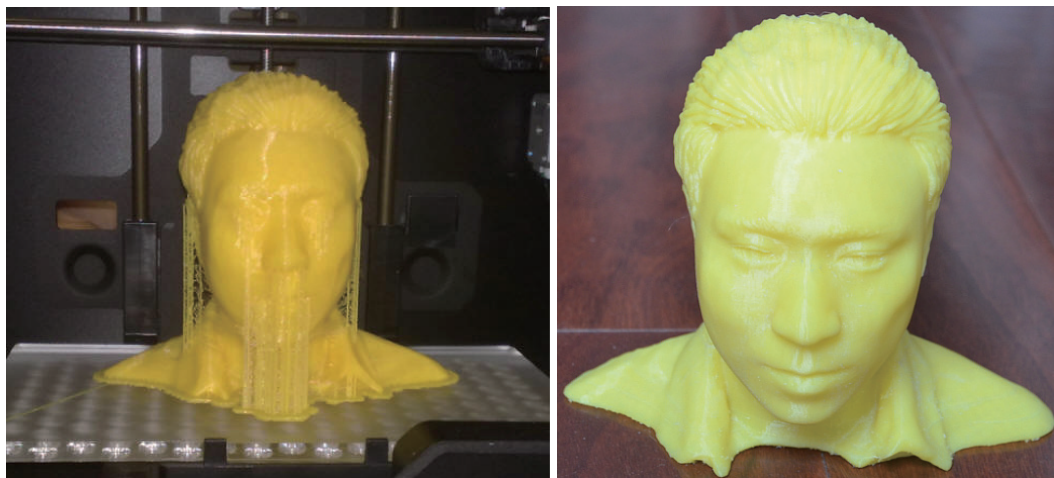


图 3-80 辅助支撑结构以及去除支撑后的最终模型

- Extruder。如果你的 MakerBot 有两个挤出头，则可选择用哪个挤出头打印你的物件。
- Cancel。单击这里取消打印。
- Make It! 发送文件到 MakerBot Replicator 2 进行打印。

设置好之后,单击 Make it! 按钮。你的 MakerBot 将对物件进行切层(Slice)并逐层打印出来。祝你打印愉快!

3.4.2 ReplicatorG 控制前台的设置：双喷头打印双色模型

类似于 Cura 和 MakeWare 的功能, ReplicatorG 也是一款前台控制软件。目前 ReplicatorG 已经停止开发维护了, 比如它的切片工具 Skeinforge 就比 MakerWare 新包含的切片工具 MakerBot Slicer 速度慢 20 倍。然而, 鉴于很多之前的 3D 打印操作指南普遍用到了 ReplicatorG, 这里我们也对它做一个简要介绍。

本节中, 我们用 ReplicatorG 实现一个特殊功能: 实现双色 3D 模型打印!

首先, 你需要先找到一个双色模型。可通过单击 ReplicatorG 的菜单 “Thingiverse” → “Dual Extrusion models!” 来在线下载一些双色模型文件。

任何一个双色模型文件都包含有两个 STL 文件, 每一个对应于一种颜色或材料。确保两个 STL 文件都下载了, 然后单击 ReplicatorG 的菜单 “GCode” → “Merge .stl for DualExtrusion” 来合并文件。

然后会看到 3 个输入框: 指定每个喷头对应的 STL 文件 (或 Gcode 代码文件) 以及合并后的 Gcode 文件名称 (必须指定 .gcode 的后缀), 如图 3-81 所示。

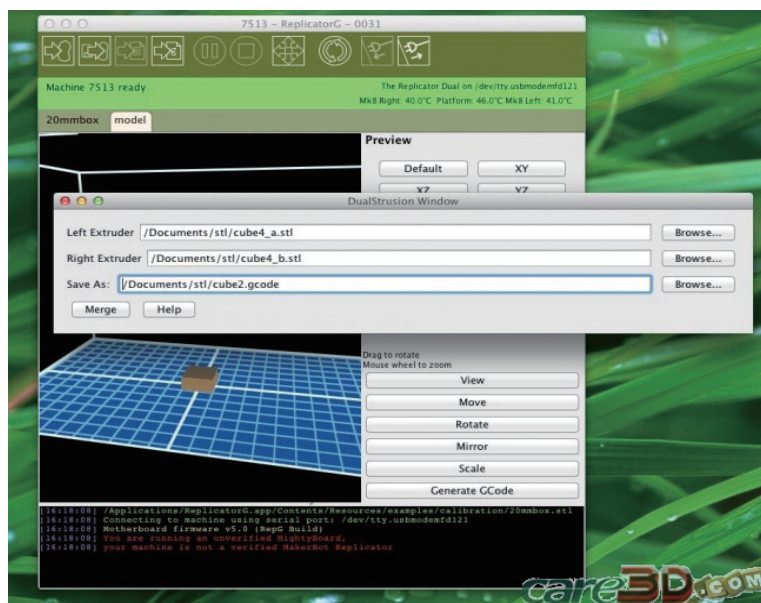


图 3-81 指定每个喷头对应的 STL 文件以及合并后的 Gcode 文件

当单击“Merge（合并）”按钮后会出现两个操作窗口：分别对应每个 STL 文件的 Gcode 配置选项。这与打印单色模型的窗口很像。确保选中“Use default start/end gcode”和“Use Print-O-Matic”两个复选框。然后分别单击各自的“Generate Gcode（生成 Gcode）”按钮，如图 3-82 所示。

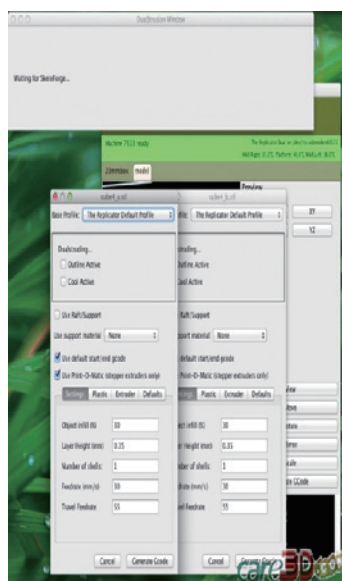


图 3-82 Generate Gcode（生成 Gcode）

ReplicatorG 将会生成两个 Gcode 代码文件（对应于两个模型），可以将它们合并，然后再生成 s3g 文件，复制到 SD 卡上就可以开始打印双色模型了，如图 3-83 所示。

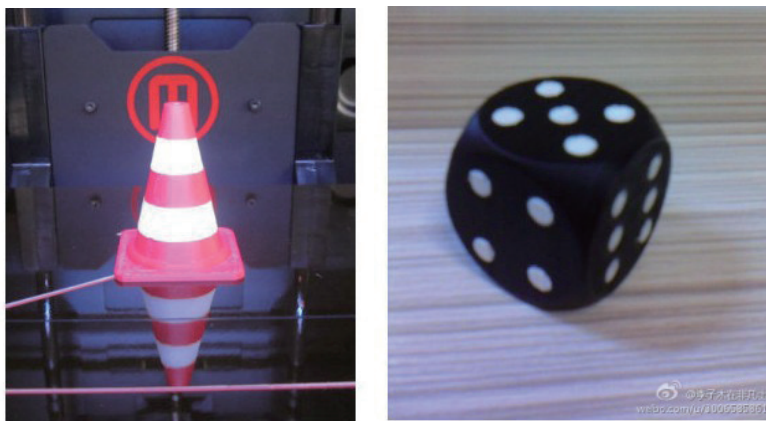


图 3-83 打印双色模型

3.4.3 MakerBot Replicator 2 打印成果实例

MakerBot Replicator 2 的打印案例如图 3-84 所示。

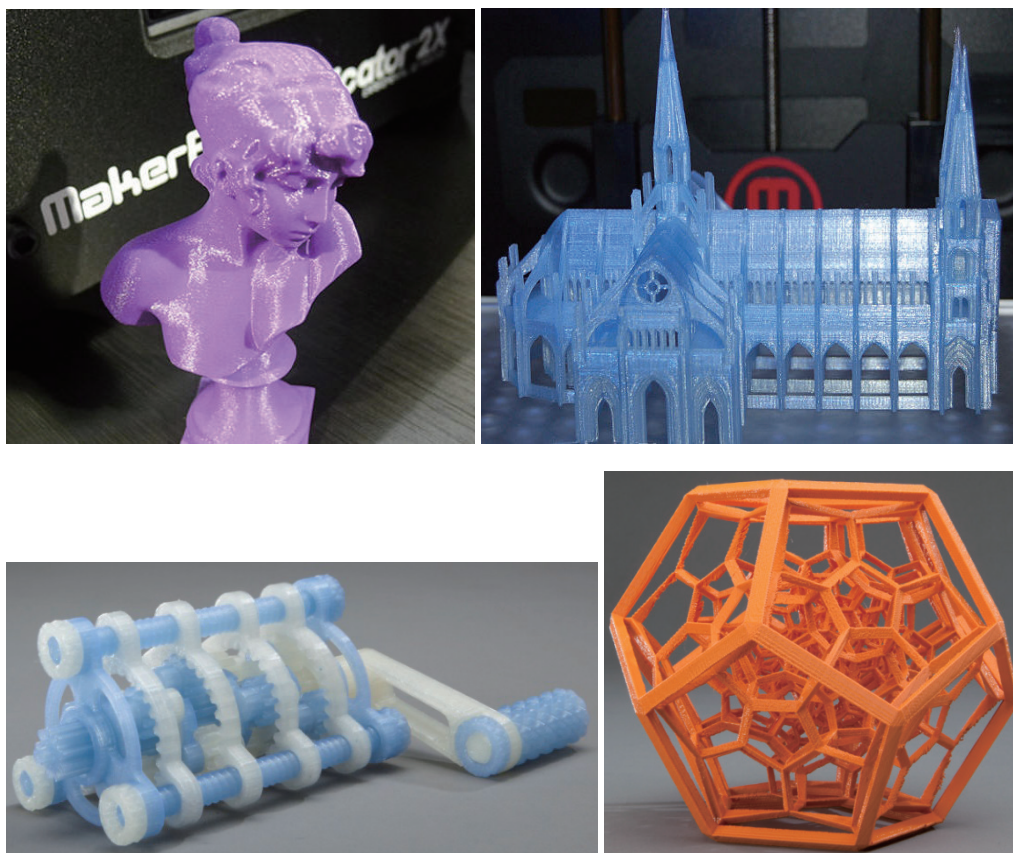


图 3-84 MakerBot Replicator 2 的打印案例（图片来源：MakerBot）

下面再提供一些细节放大图，如图 3-85 所示。



图 3-85 MakerBot Replicator 2 的更多打印案例（细节放大图）

3.5 3D打印疑问与故障排解小贴士

下面，我们对 3D 打印中经常会遇到的一些问题做一个集中解答。值得注意的是，目前 3D 打印机有诸多型号，所以这里只能列举一些共性问题，在实际操作中请务必参考厂商提供的详细文档和教程，此外还要在网络上搜索相关资料^{[52][53][55]}。

3.5.1 模型的水密性（Watertight）

一般而言，3D 模型文件需要水密化后才可以进行三维打印。水密性也可理解为密封性，也就是“不漏水的”，把水充满模型中间而不会漏出来，这就要求模型上不能有孔洞。你可能会感到惊讶，很多 3D 模型都会存在一些难以察觉的小孔。如果你的眼力不够“尖”，可使用软件（如 AccuTrans）自动查找这些小孔。

3.5.2 模型必须为流形（Manifold）

3D 模型必须为流形。通俗地说，如果一个网格模型中存在多个（3 个或以上）面共用一条边，那么它就是非流形的（Non-Manifold），因为这个局部区域由于自相交而无法摊开展平为一个平面了。请看如图 3-86 所示这个 4 个面共享一条边的非流形例子。

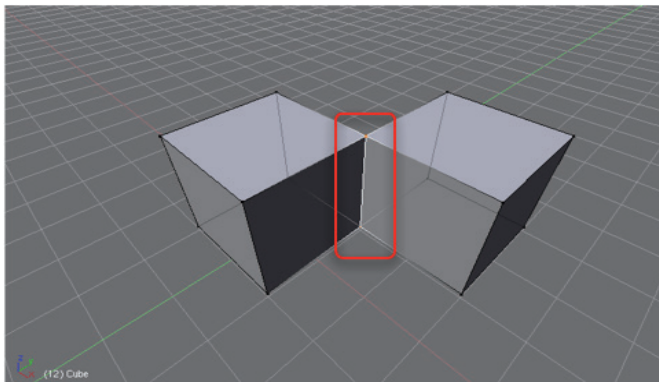


图 3-86 非流形模型的例子（4 个面共享一条边）



提示:所谓**流形 (Manifold)**,是局部具有欧几里得空间性质的空间。严格的数学定义为:一个 n 维的流形是一个连通的 Hausdorff 空间 M , 即, 对于 M 中的任何一点 p , 都存在有一个邻域 $U \subset M$, 其同胚于 R^n 欧几里得空间的开子集。

这个深奥且拗口的数学定义可用通俗的话来理解。如图 3-87 所示的地球球面就是一个二维流形。因此,对于球面上的一个曲面三角形(左侧),可以摊开展成(即流动变形成)一个二维欧几里得空间上的平面三角形(右侧)。此外,因为地球实在太大,我们往往把地球上的一块足够小的(曲面)局部区域当作平面来丈量,而不用担心会引起大的误差。比如,你要丈量学校操场的面积,根本不用把它认为是地球上的一块曲面,而直接看作一块平面即可。所以,光滑流形其足够小的结构是“硬”的(如可以固定丈量),而整体结构则是“柔软”的(可流动变形)。**流形 (Manifold)**可看作是很多(Many)曲面片的叠加(Fold),比如整个地球的地图册就是由各个地区的地图页合订而成的,而相邻地区的地图页之间含有重叠区域,以便建立彼此之间的联系,这样我们才能通过翻看一页一页的局部地图得出整张世界地图。

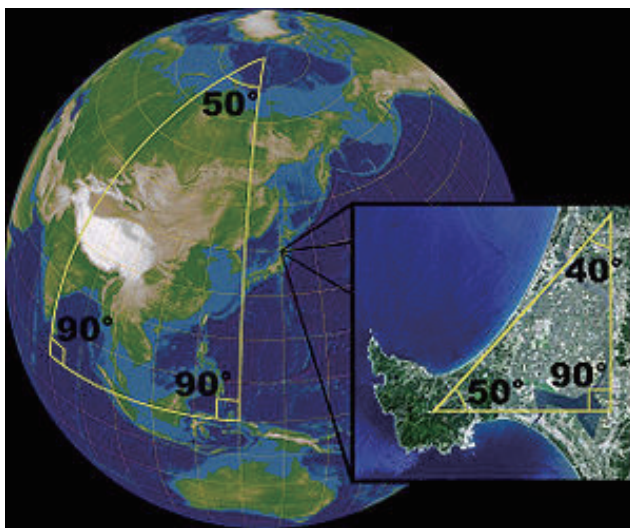


图 3-87 球面为二维的流形, 因为可由一群二维的平面图形来叠加表示 (图片来源: 维基百科)

3.5.3 切片 (Slice) 与横切面

在 3D 打印前,软件会将 3D 数字模型“切片”(Slice),生成由一系列横切面 (Cross-Sections) 组成的 Gcode 数字化文件。切面的厚度实际上就是打印机支持的单层厚度(比如 0.1mm)。然后,3D 打印机的工作就是:读取 Gcode 文件,不断地将这些横切面层层堆积,直到将 3D 实体打印完成。常见的切片软件引擎(又名 Gcode 生成器、POST 软件) 有: CuraEngine (Ultimaker)、Skeinforge、Slic3r、MakerBot Slicer (MakerBot)、SFACT、KISSlicer 等。

3.5.4 层厚度 (Layer Thickness)

3D 打印工艺都有各自的规格限制,其中最重要的一项就是机器所打印的每层的厚度 (Layer Thickness)。如果你的设计中存在精细到 0.01mm 的细节,而打印机的精度只有 0.1mm,那么你就只能跟精心设计的细节说拜拜了,因为打印机会自动忽略它,所以针对应用要求选择相应精度 (层厚度) 的 3D 打印机尤为重要。

3.5.5 支撑材料 (Support Material)

对于 3D 形状上的悬垂或中空结构,一般都会使用支撑材料来保证模型不会在打印过程中坍塌掉。支撑材料一般都比较好去除,成本通常比模型材料便宜一些。如果图省事的话,你也可以直接采用模型材料来作为支撑材料。



提示: 记住 45° 法则。任何超过 45° 的突出物都需要额外的支撑材料或是高明的建模技巧来完成模型打印。善用“老鼠耳朵 (Mouse Ear)”,一种圆盘状或是圆锥状的底座,把它们设计到你的模型之中,尽量避免使用内建的打印底座 (Raft), 它会拖累你的打印速度或难以去除并且损坏模型。

3.5.6 如何开始打印

第一次使用 3D 打印机时,可按照如下步骤操作。

1. 收到打印机后,拆箱取出打印机和免费赠送的配件。
2. 在计算机里安装好附赠 SD 卡中的打印机软件 Cura、Repetier Host 等。
3. 使用 100~240V 电压,通过电源适配器给打印机接上电源,调平打印平板。
4. 给打印机上料 (ABS、PLA 等线材)。
5. 准备好 3D 模型文件 (一般为 STL 格式), 并通过 Cura 软件转化为打印机可识别的 Gcode 文件。
6. 将装有 Gcode 文件的 SD 卡插入打印机 SD 卡槽,通过菜单选中该 Gcode 文件。
7. 等待打印机自动打印。

在打印机处于正常使用状态而且打印材料充足的情况下,直接操作 5、6、7 这 3 个步骤就能打印了。

3.5.7 如何调平打印平台（粗调和精调）

保持打印平台（即贴有蓝色胶布的那块打印平板，蓝色胶布的作用是使打印件更容易剥离打印平台）的水平对于打印质量非常重要。在 3D 打印机的使用过程中，一般来说，可以跳过粗调直接进入精调。然而，如果打印机长时间未使用，粗调步骤还是必要的。粗调的目的是保证平台在每次取放后，平台和喷头保持 0.5mm 左右的合理距离。

1. 转动 Z 轴螺纹丝杆，直到喷头几乎触及平台。
2. 移动打印头到左前角。
3. 调节平台上的调节螺丝，直到喷头接触平台。
4. 把平台右前角往下调整，直到你能在不挂伤平台的情况下移动喷头到平台右侧。
5. 然后上调平台右前角，直到它接触到喷头。
6. 重复此过程来调整 4 个角落的调整螺丝。

精调是指通过运行厂商自带的调平软件，进一步缩小喷头与平台的距离（大约一张 A4 纸的厚度）。

A) 调平软件运行后，喷头会依次移向平台四角，移动到每一个角时，喷头都会暂停。

B) 喷头每次暂停时，在喷头下插入一张白纸（A4 纸）。如果喷头与平台间距离大于白纸厚度，则逆时针拧松平台上该角落处的螺丝，减小平台和喷头的距离，并左右移动白纸，直到能感觉到白纸与喷头之间的强烈的摩擦感，且喷头在白纸上留下明显而不是特别重的刮痕。反之则顺时针拧紧平台上该角落的螺丝。

C) 以上过程重复进行两遍之后，喷头开始加热到 220℃，随后吐丝打印样例方框。在此过程中，随时调节平台四角的螺丝，如果样例方框线粗细均匀，无拉丝、断裂现象，表明平台已经调平。

3.5.8 如何更换耗材（上料、退料）

换塑料丝的最简单方法是使用打印程序内置的脚本“Utilities”→“Filament Options”，有装载或卸载挤压机塑料丝的选项，按照说明操作即可。

你也可以用软件（如 ReplicatorG）手动装卸 ABS 塑料丝。首先将挤压机工作温度调到 225℃，然后设置挤压机的转速为 3.0r/min，让电机反向转动后就可以取出塑料丝，基本上不到 30s 就可以完成，相反，只要正向转动就可将塑料丝重新装入。

3.5.9 我装不了塑料丝

装塑料丝的时候用大点儿力，特别是对于新的打印机可能需要用更大的力气。如果仍装不进，则找剪刀剪掉塑料丝头部的一小段，并剪成斜角，然后用力抓紧它垂直地插入挤压机的顶部对应位置，也可以用钳子或其他东西夹紧，插的过程中不要倾斜。

注意，装塑料丝是不会弄坏打印机的，尽可能用大点儿力气。当发现塑料丝开始进入挤压机时，继续保持向下压 10s。

3.5.10 我取不出塑料丝导管

拉导管不能靠蛮力，你需要先按住挤压机顶部的灰色塑料环。塑料环的设计就像一把锁，不往下很难打开它，所以必须用手指按住这个环，一次不行就换换按其他的位置，不过要注意轻点。

3.5.11 为什么我的送料机挖坑，但就是不吐丝

挖坑说明送料机没问题。需要做的是：

- 检查一下挤出头，试下手动推料是否出料正常。
- 检查一下配置的温度，是否合适。
- 检查一下软件参数，是否改动了什么导致这个问题。
- 检查一下回抽设置，是否回抽得太厉害了，导致料出来后进不去。

3.5.12 喷头堵塞，如何处理

可做如下操作。

- 找根针捅捅，加热的时候捅。
- 拆喷头，清理喷头里面残留的耗材。
- PLA 堵头，可以先将温度升高至 240℃，再打印，或许可以融化里面的残留物。

3.5.13 挤出的料无法粘牢打印平台

挤出的料无法粘牢平台，模型的第一层无法和打印平台粘牢。请仔细检查：

- 亚克力平台在使用前是否贴上了胶带。
- 挤出头和平台间距要有一张纸的厚度。
- 是否出料不足导致？正常情况打印时，喷头的料可以自然流淌。
- 温度设置是否合适。

3.5.14 打印出的东西粘不牢平台

别担心，这一问题大家都会遇到。解决方法是，将打印平台清理干净或者将它稍稍调高一点，用一块不掉毛的绒布加上一点点酒精或者一些丙酮指甲油清洗剂将平台表面抹干净。丙酮可以在五金店找到，使用它前要认真阅读说明书，因为丙酮有毒性。

以上方法都不行的话试试旋转打印平台底部的旋钮，调整平台高度，让平台贴近喷头。如果还是粘不住，运行打印机的校准程序，在“Utilities”→“Level Build Plate”打开。这次用一张更薄的纸片放在平台上进行校准，目的是调整平台让喷头更靠近这张纸片，这步可能要多试几次。

还有个选择是用打印底座（Raft）的方式进行打印，这种方法可让你的打印物体更加容易粘上平台。

3.5.15 用辅助盘（Helper Disks）解决翘边问题

在用 ABS 材料打印 3D 模型时，有时会遇到一个头疼的问题：翘边问题。在打印大尺寸或者底部面积较大的模型时，翘边会变得更为突出。



提示：3D 打印翘边的根本原因是：模型的底部边缘与基底黏得不牢靠，温度的快速降低会导致材料收缩，因此出现了翘边问题（如图 3-88 所示）。具体影响因素有：平台底盘预热不均、打印速度较慢、ABS 打印材料的弹性和收缩度不够等。

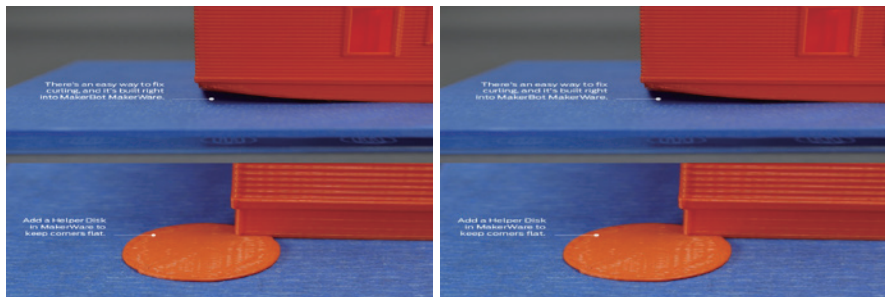


图 3-88 翘边问题以及用辅助盘来解决（图片来源：Thingiverse）

可以使用辅助盘（Helper Disks）来避免模型翘边问题。打印后，辅助盘也很容易拆除。

下面介绍一下操作步骤。打开 MakerWare 软件，导入 3D 模型。如图 3-89 所示的这个房屋模型有很多的边角，打印时很容易产生翘边。在“File”菜单中，打开“Examples”菜单，选择符合大小的“Helper Disks”文件。然后把辅助盘放置到房屋的角上，只需被压住一部分即可。

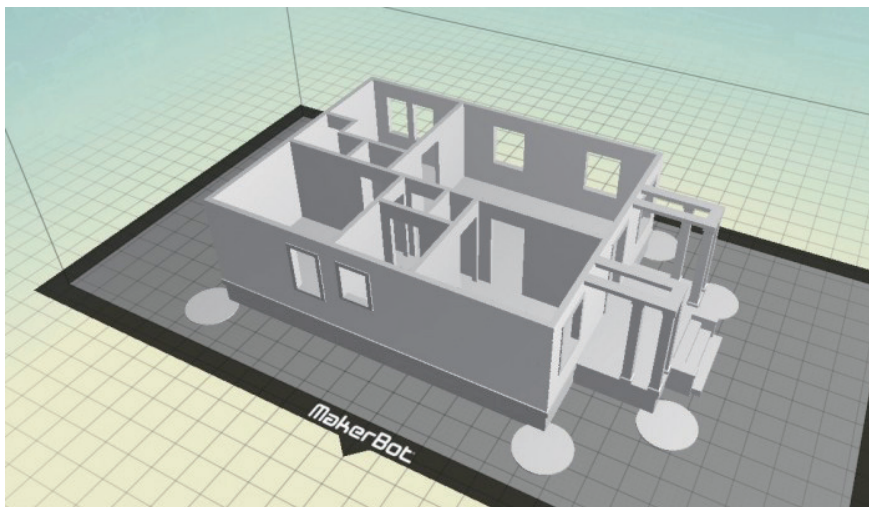


图 3-89 Rodessa House 模型的第一层，有很多边角，打印时容易翘边

最后单击“Make It”按钮，开始生成切片打印模型。当模型打印完成后，这些辅助盘很容易取下。通常情况下，直接撕下即可，也可以用类似剪钳的工具取下这些小盘。

使用辅助盘来解决 ABS 材料的翘边问题，问题虽然可以得到缓解，但终究还是比较烦琐。

目前，PLA 材料越来越多地应用到 3D 打印。与 ABS 材料相比，PLA 在硬度、弹性与收缩性上有明显优势，同时抗翘边的性能更好。有关 PLA 材料的详细介绍，请参见第 2 章 2.3 节“塑料还是石膏？3D 打印机的各种耗材”。

3.5.16 喷头位置偏移，挤出头坐标异常

打印过程出现电机失步，喷头位置偏移，挤出头坐标出现异常，可能的原因有：

- 同步轮没上紧。
- 光轴污物太多，导致阻力太大，电机失步。
- 挤出头上的白色透明管弹性过大。

确定相应的原因后，针对性地解决即可。

3.5.17 为什么打印的圆是椭圆

请仔细调整 X/Y 轴的正交度。移动 X 轴紧贴框架，校准 X 轴两端与框架的间距，使其一致来保持 X 轴与框架对应一侧的平行度。同样地，对 Y 轴做一次校准，保持 Y 轴与框架对应一侧的平行度。若 X 轴、Y 轴分别与框架两侧平行，则相互正交。

3.5.18 电机不转，像得了帕金森症抖个不停

电机的 4 针线没插牢固。即便是白色塑料接头已经牢固，也要检查里面的针头是否牢固。

3.5.19 为需要连接的零件选择合适的容许公差

为拥有多个连接处的模型设计你觉得合适的容许公差。计算公差的技巧是：在需要紧密接合的地方（压合或联结物件）预留 0.2mm 的宽度；给较宽松的地方（枢纽或是箱子的盖子）预留 0.4mm 的宽度。

在要求精度的模型上不要使用过多的外壳（Shell），像是对于一些印有微小文字的模型来说，多余的外壳会让这些精细处模糊掉。

此外，你需要懂得调整打印方向以求最佳精度，永远以可行的最佳分辨率方向来作为你的模型打印方向。如果有需要，可以将模型切成好几个区块来打印，然后再重新组装。对于使用 FDM 技术的打印机来说，你只能控制 Z 轴方向的精度，因为 XY 轴的精度已经被线宽决定了，如果你的模型有一些精细的设计，确认一下模型的打印方向是否有能力打印出那些精细的特征。

3.5.20 如何让模型表面更光滑

一直以来，采用熔融沉积成型（FDM）3D 打印的产品，无论是使用工业级还是桌面级 3D 打印机，都有一个很难解决的问题：打印出来的产品都会显示出一些层效应（Layered Effect）。虽然我们可以用砂纸或锉刀进行打磨，但这属于材料去除工艺，如果分寸拿捏不当，则可能反而对精度和细节有损害。

最近重庆科技学院的一名本科学生研发出低价 3D 打印抛光机，这种抛光机不是采用传统的

去除材料的方法，而是采用他们称之为“材料转移技术”的方法达到抛光目的，将零件表面突出部分的材料转移到凹槽部分，对零件表面的精度影响非常小。抛光过程中不产生零件的废料，零件的分量也不会改变。

最重要的是他们不使用丙酮、丁酮、氯仿、四氢呋喃等剧毒物质抛光，而是使用一种自主研发的环保耗材。他们所使用的打印机是 5 000 元左右的廉价桌面级 3D 打印机。从图 3-90 中可以看到，打印出来的效果一般般，但经过抛光后，人的面部五官不仅光滑且轮廓分明。



图 3-90 对模型表面进行抛光处理（图片来源：重庆科技学院）

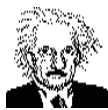
3.5.21 我的打印机需要日常维护吗

当然需要！个人 3D 打印机通常会提供一个可自加热的打印托盘（也被称为**热床**），要么是标配要么是选配。热床用来防止打印物品在冷却时变形或开裂，同时保证 ABS 材料的物品底部牢牢地粘在打印托盘上。为了保证良好的导热性，以及表面的平滑和水平，热床的上面一层通常是玻璃或铝板，玻璃的光滑性更好，而铝板的导热性更好。此外为了在打印过程中增强耗材与平台的附着力，并减少打印时的翘边和变形，在托盘表面通常会贴一种价格便宜、可定期更换的胶带，胶带的材质包括 Kapton 类的聚酰亚胺胶带或 PET 类的聚酯胶带，甚至是各种蓝色的纸胶带。

因此，你需要经常更换 Kapton 防热胶带，此外还需要给打印机上润滑剂，重新拧紧螺栓等。

3.5.22 异常情况如何中断打印

用户如果由于某种原因中途不想再打印，可以按下 LCD 显示屏的旋钮，进入 SD 卡主操作界面。选择“Stop print”来取消打印任务。



注意：即使按下“Stop print”，喷头的温度也不会降下来，仍然保持在设定温度，如 220℃，所以操作的时候要小心高温。如此设置主要是为了方便下一次的打印，这样你就不用浪费时间等待喷头重新加热了。

3.5.23 如何将金属零件放入我的 3D 塑料模型中

目前大家买得起的一般都是以塑料为耗材的个人桌面级 3D 打印机，因为金属打印机现在仍

很昂贵。难道我们就只能打印一些塑料模型吗？非也！这里介绍一种叫“**包埋**”的方法，几乎在所有的 3D 打印机上都能用。原理非常简单：你只需中途将打印暂停，将所要包埋的金属零件（比如螺母、螺栓、电动马达）放入尚未封合的打印物中，然后再继续打印就可以了。实现的关键在于：零件要设计成可容纳螺母大小的镂空结构，并把镂空结构稍微弄大点以便放入。然后观察打印过程，在打印完螺母腔的顶端那一层之后将打印机暂停。因为眼睛有时很难判断是否打印到最后一层，因此可在设计时增加一个参照标线来帮助判断。

3.5.24 用 CNC Simulator 进行打印模拟和打印预览

对于复杂的 3D 模型，第一次打印很可能会失败。为了防患于未然，我们可利用 CNC Simulator 模拟器对 3D 打印进行模拟，预览一下是否能打印出自己想要的效果，并对可能的打印失败进行纠错。CNC Simulator 模拟器运算速度很快，允许多种类别的打印前调试，例如，从不同角度查看打印过程、选用打印材料（从塑料到金属），以及改变打印材料的颜色等。

3.5.25 打印失败后是什么样子

前面我们展示的都是打印成功的案例。而实际上，对于新手来说，一开始的打印往往是失败的。虽不像“一将功成万骨枯”那么夸张，但“一将功成十骨枯”还是非常可能的。下面，我们就对一些打印失败的案例进行展示，并解释失败的原因。

如图 3-91 所示的打印失败案例是由 X 轴和 Y 轴马达滑步造成的。



图 3-91 X 轴和 Y 轴马达滑步造成的打印失败（图片来源：Flickr）

本来要打印迪斯尼雕像，但 Gecko 步进电机驱动错误，于是造成如图 3-92 所示的结果。

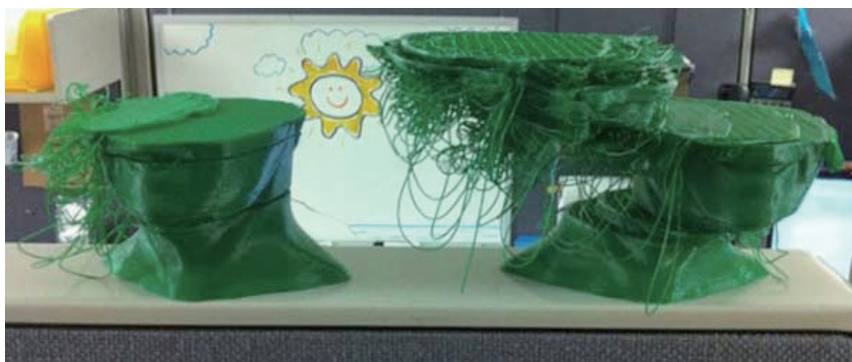


图 3-92 Gecko 步进电机驱动错误造成的打印失败

本想打个方块儿，但由于高度设置错误，结果造成挤出错误变成了一团“方便面”，如图 3-93 所示。

打印机校对出错，后果果然很严重，如图 3-94 所示。



图 3-93 高度设置错误造成的打印失败

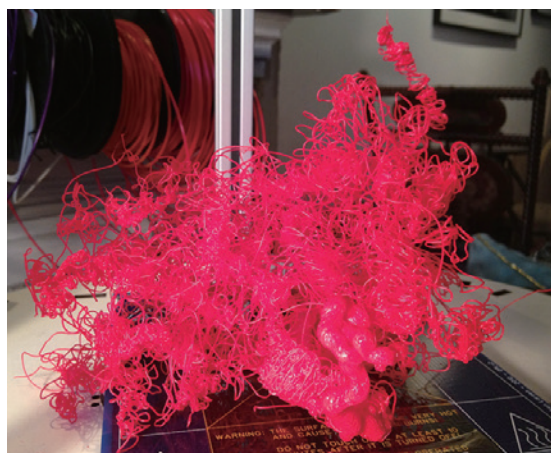


图 3-94 打印机校对出错造成的打印失败

第4章

3D智能数字化：3D打印的孪生兄弟

《道德真经集解》(宋代·董思靖)有云：“虚乃实之根”，实从虚生；实乃虚之体，虚以实成型。这一句话对虚与实的辩证哲学关系做了精彩的论述。一千年后的今天，3D智能数字化与3D打印实现了从真实物体到虚拟三维，然后又从虚拟三维再造实体的轮回转换。正可谓：虚即是实，实即是虚！

从本章开始将重点介绍3D智能数字化。这不仅是科学界的一个热门研究领域，而且与我们的现代日常生活密切相关。比如，家里客厅摆放的是3D数字电视、去电影院看的好莱坞大片是3D电影，智能手机里玩的是3D游戏，等等。如果说数字化是信息社会的媒介和载体，那么智能化则是核心和灵魂，使得数字内容有了灵性、不再空洞乏味。

因此，3D智能数字化不仅仅是对真实世界的描绘和刻画，而且通过智能化设计工作，还完全可以创造出一个更加美好的世界来。以电影《阿凡达》为例，很多美轮美奂的场景都无法从现实中直接拍摄，而通过数字化的艺术设计，再使用3D打印机直接打印出来，这就获得了超越现实的逼真效果。3D智能数字化与3D打印的完美结合，实现了用“虚拟”再造“现实”的崭新境界。

4.1 不以规矩，不成方圆——STL数字标准文件解析

STL (STereo Lithography 的缩写) 文件格式由3D Systems公司的创始人查尔斯·W·哈尔 (Charles W. Hull) 于1988年发明，当时主要针对Stereo Lithograph (光固化立体成型) 工艺，现已成为全世界CAD/CAM系统接口文件格式的工业标准，是3D打印机支持的最常见3D文件格式。

当你将3D模型保存为STL文件后，物体的表面轮廓形状会被转换成三角形面片网格，如图4-1所示。每个三角形面片的描述包括三角形各个顶点的三维坐标及三角形面片的法向量。

STL文件格式具有简单清晰、易于理解、容易生成及易于分割等优点。STL文件分层处理只涉及平面与一次曲线求交，分层算法相对简单。此外，还可以很方便地控制STL模型的输出精度。

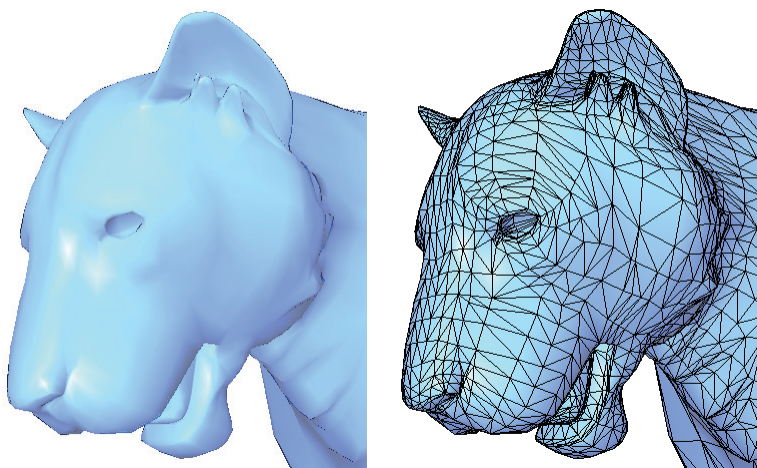


图 4-1 三角形面片网格

STL 文件有两种：一种是 ASCII 文本格式（可读性好，可直接阅读），另一种是二进制格式（占用磁盘空间小，为 ASCII 文本格式的 1/6 左右，但可读性差）。

STL 的 ASCII 文本格式

ASCII 文本格式的 STL 文件逐行给出三角形面片的几何信息，每一行以 1 个或 2 个关键字开头。整个 STL 文件的首行给出了文件路径及文件名。STL 三维模型由一系列三角形面片构成。每一个三角形面片（facet）由 7 行数据组成，facet normal 是三角形面片的法线方向，outer loop 说明随后的 3 行数据分别是三角形面片的 3 个顶点坐标，这 3 个顶点沿指向模型外部的法线方向逆时针排列（遵循右手法则）。

ASCII 文本格式的 STL 文件结构如下。

```
solid objectname           // 物体名
facet normal x y z         // 三角形面片法向量的 3 个分量值
    outer loop
        vertex x y z       // 三角形面片第 1 个顶点坐标
        vertex x y z       // 三角形面片第 2 个顶点坐标
        vertex x y z       // 三角形面片第 3 个顶点坐标
    endloop
endfacet                   // 完成一个三角形面片定义
.....
endsolid objectname       // 整个 STL 文件定义结束
```

STL 的二进制格式

二进制格式的 STL 文件结构如下。

```
UINT8[80]                  // 文件头信息
UINT32                     // 面片数目
foreach triangle
    REAL32[3]               // 某个面片的法向量
    REAL32[3]               // 某个面片第 1 个顶点
```



```
REAL32[3]          // 某个面片第 2 个顶点
REAL32[3]          // 某个面片第 3 个顶点
UINT16             // 某个面片的 16 位整数型属性字，一般为 0，无特别含义
end
```

可以看出，无论 ASCII 文本格式，还是二进制格式，STL 文件格式都非常简单、一目了然。

此外，实际应用中 STL 模型数据是有要求的，要经过检验才能使用。检验主要包括两方面的内容：STL 模型数据的有效性和 STL 模型封闭性检查。有效性包括检查模型是否存在裂隙、孤立边等几何缺陷；封闭性则要求所有 STL 三角形围成一个内外封闭的几何体。

截至目前，国内外已有很多研究人员针对 STL 模型数据处理做了大量卓有成效的研究工作，这些工作主要集中在：

- STL 文件的错误检测与修复。
- STL 文件模型的拓扑重建。
- STL 文件模型的分割。
- STL 模型的分层处理（等层厚及变层厚）。
- 基于 STL 文件的三维模型分层方向优化。
- 基于 STL 文件的支撑生成与优化。
- 基于 STL 文件的层片扫描路径的生成及优化。

然而，STL 文件格式还是显得有点过于简单了，其只能描述三维物体的表面几何信息，不支持颜色、材质等信息（而另一种常见的 PLY 文件格式就可以支持彩色纹理），也无法表达形状内部的中空结构。因此，2011 年 7 月，美国材料与实验学会（ASTM）发布了一种基于 XML（可扩展标记语言）的增材制造文件 AMF（Additive Manufacturing File）格式。相比于 STL 文件格式，AMF 格式可处理不同类型的材料、不同颜色（包括渐变色，如图 4-2 所示）、曲面三角形以及复杂的内部中空结构（这正是 3D 打印最重要的优势之一）。与 STL 采用的平面三角形相比，曲面三角形可以更准确、更简洁地描述曲面。

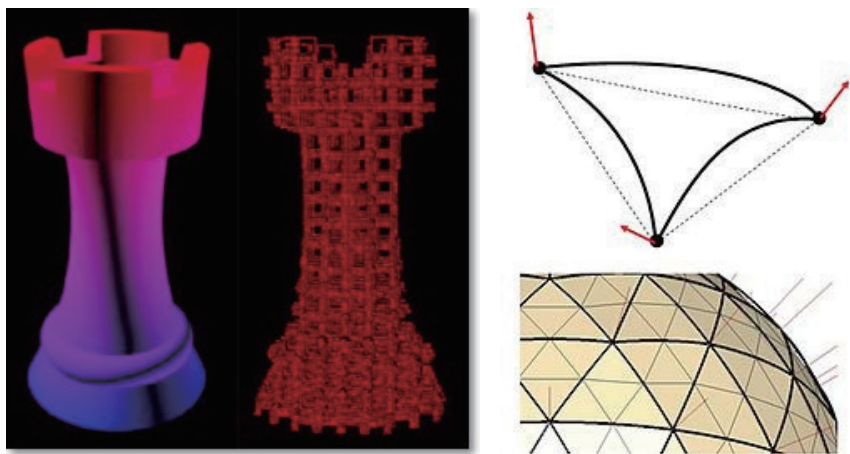


图 4-2 AMF 文件格式：可描述渐变颜色以及曲面三角形（图片来源：ASTM）

AMF 格式虽好，可谓“青出于蓝而胜于蓝”，然而因为 STL 格式已经根深蒂固了，所以 AMF 想要“上位”，乃至被广大 3D 打印机厂商普遍采用，估计尚需等待一段时日。

4.2 3D智能数字化设计技术

3D 智能数字化技术主要分为两大类：3D 智能数字化设计和 3D 智能数字化扫描。在本节中我们将介绍 3D 智能数字化设计，来从无到有地设计 3D 数字化产品。

3D 智能数字化设计与传统的 CAD 技术关系密切，是其不断发展所抵达的最新阶段。所谓 CAD (Computer Aided Design, 计算机辅助设计)，是指利用计算机软件制作并模拟实物设计，展现新开发产品的外形、结构、色彩、质感等特色的过程。



提示：除了 CAD，还有 CAM (Computer Aided Manufacturing, 计算机辅助制造，利用数控机床控制刀具运动，完成零件制造)、CAE (Computer Aided Engineering, 计算机辅助工程分析，利用有限元计算方法分析产品的结构强度、热传导等性能)、CAPP (Computer Aided Process Planning, 计算机辅助工艺过程设计，利用计算机技术设计产品的加工工艺和步骤)、CAI (Computer Aided Instruction, 计算机辅助教学) 等。CAD、CAE、CAM、CAPP、CAI 等统称为 CAX。

随着社会对数字化生存的依赖日益加强，仅靠 CAD 技术已难以应付各行各业的需求，因此 3D 计算机图形学、计算机视觉、模式识别与智能系统、机器学习等其他交叉学科已开始融入进来，且越来越有融为一体的趋势。在本书中，我们将这种新的数字化技术统称为 3D 智能数字化技术。

4.2.1 “所想即所得”：3D 设计的新境界

3D 设计分为两大类：实体建模和曲面建模。**实体建模 (Solid Modeling)** 主要面向工业设计和制造领域，如将一个圆柱体零件和一个正方体零件合并在一起，或在一个球体零件上钻一个方孔。而**曲面建模 (Surface Modeling)**，正如字面意思所揭示的，只考虑形状的表面（内部可认为是个空壳），主要面向影视动漫、游戏娱乐领域（这些领域只要求形状的外表看着逼真就行，内部是空的也没有关系）。

实体建模一般用来设计规则的几何形状，对于不规则的几何形状则有些力不从心。而 3D 打印的特色就在于制造那些独特的不规则形状物体。因此，实体建模工具以后可能会慢慢过时。另一方面，曲面建模可以胜任复杂、精细的不规则形状（这从好莱坞大片中细节越来越逼真的怪兽、变形金刚就可窥豹一斑），然而它的缺点是形状内部是空的，无法描述形状内部的“满园春光”或“内藏乾坤”，比如复杂精巧的内嵌结构。

目前大多数 3D 设计软件都既可以实体建模，也可以做曲面建模。因此，最简单的建模方法就是手工使用这些 3D 设计工具，如 SolidWorks、AutoCAD、3DS Max、Maya、Rhino3D、ZBrush 等，像在沙滩上玩沙雕一样堆积、组合、掏空实体，或像裁缝一样将一块曲面反复裁剪、拉伸、变形成最终的形状。



提示：曲面有很多种表示方法，常见的有3种：NURBS 曲面、多边形曲面和细分曲面。
NURBS 是非均匀有理 B 样条曲线（Non-Uniform Rational B-Splines）的缩写，它的优点是只需操纵少数的样条控制点就可以拟合各种光滑的形状，如图 4-3 所示。

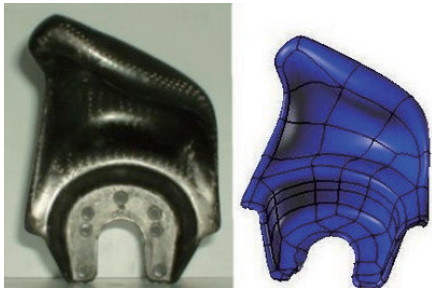


图 4-3 用 NURBS 曲面拟合零件形状(图片来源：Rhino)

而**多边形（Polygon）**曲面一般为**三角形网格（Triangle Mesh）**，使形状由许多个三角形面片来表示，可表现复杂形状的精细细节。如图 4-4 所示是 NURBS 曲面转换成多边形曲面的例子。此外在工业界，人们更偏向于使用**四边形网格（Quad Mesh）**，而不是三角形网格，这是因为四边形网格的边更能反映物体表面的流线方向（Edge Flow），从而便于建模工具进行细节的生成和编辑。

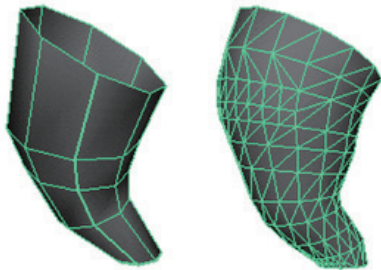


图 4-4 NURBS 曲面（左）转换成多边形曲面（右）(图片来源：Autodesk)

细分（Subdivision）曲面，又被称为子分曲面。你只需制作一个粗糙的控制网格，然后指定一个细分规则（如 Catmull-Clark 规则、Loop 规则），就可以自动将粗糙网格不断细化成任意光滑的曲面。好莱坞的皮克斯工作室（Pixar，现属于迪士尼）的创始人之一 Edwin Catmull 就是细分曲面的主要发明者，因此不难理解 Pixar 的 3D 动画片大量采用了细分曲面来表现圆润平滑的形状边缘，如图 4-5 所示。

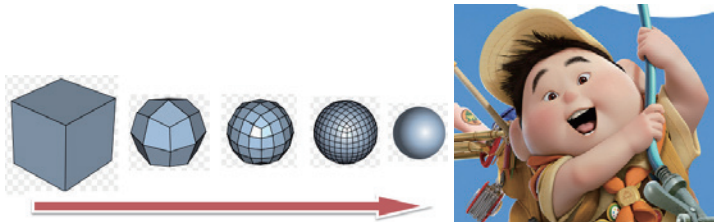


图 4-5 细分曲面可表现圆润平滑的细节(图片来源：Pixar)

手工建模是一件比较烦琐、费时的的工作。比如设计一把椅子，设计完成后，如果发现中间的某个地方尺寸短了，那我们不光要修改这个地方，还要修改与之相连的两端，否则这把椅子就合拢不上了。参数化建模解决了这个问题。所谓**参数化建模**（**Parametric Modeling**，也被称为基于特征的建模），就是将原有设计中的某些尺寸特征，如形状、定位或装配尺寸，设置为参数变量（所谓变量，也即不是定死的，而是可灵活调整的）。如果修改这些变量的值，计算机就会自动变动其他相关的尺寸，由此得到不同大小和形状的零件模型。参数化设计的本质是在可变参数的作用下，系统能够自动维护所有的不变参数（如椅子腿长必须为 0.5m，椅子后背的长宽比必须为 2:1 等），以保持形状的固有特征。有了参数化设计，我们只需简单指定长、宽、高这 3 个参数，就能快速获得一大堆定制的茶杯形状模型，而无须费时费力地对每个茶杯的几何细节（壁厚、手柄、底座等）尺寸逐一做手工更改。

参数化设计固然灵活，但需要手工设置特征变量和约束关系，所以也不是一件特别轻松愉快的活儿。于是，新一代的**直接建模**（**Direct Modeling**），奉天承运、应运而生了。所谓直接建模，就是不管原有模型是有特征还是无特征的（比如从其他 CAD 系统读入的非参数化模型），都可以直接进行后续模型的创建，而无须关注模型的建立过程，也无须维护模型树和历史树（以创建的时间顺序列出零件的参数特征和约束关系）。这样就使得我们可以在一个自由的 3D 设计环境下工作，以直观的方式对模型直接进行编辑。直接建模使我们可以在自然流畅地直接在模型上动态操作、实时预览，所有交互都直接对模型本身进行，修改模型也变得非常简单，如图 4-6 所示。

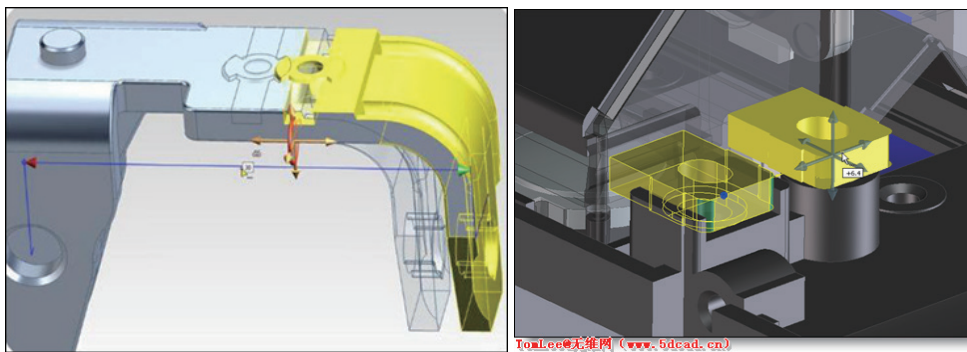


图 4-6 直接建模（动态尺寸编辑与特征位置编辑）（图片来源：PTC）

人类追求机器智能化的努力是无止境的（当然，原始的初衷大部分是为了让自己得以偷闲）。于是，更加智能化一点的**编程式设计**出现了。计算机把形状的设计过程描述成一系列有特定顺序的操作步骤，这有点像按照食谱而不是最终的外观来制作蛋糕。程式化智能设计可以轻易地在这个蛋糕上绘制几百万个规则的精美图案，而这对于手工设计来说犹如噩梦（设计师可能会被活活累瘫）。

为了生成更加丰富多变的个性图案，还可采用复杂的生长式智能系统。按照一套既定的生长规则加上随机扰动，随着时间的推移发展，将一颗种子形状不断迭代分裂，不断长出新的枝叶，最终生长成独一无二的特定形状。这种方式的建模方法，专业术语叫作**过程建模**（**Procedural Modeling**），也就是说形状的建立有一个生长的过程，代表性的方法有用于植物建模的 L 系统（L-Systems，如图 4-7 所示）以及大名鼎鼎的分形（Fractal）等。

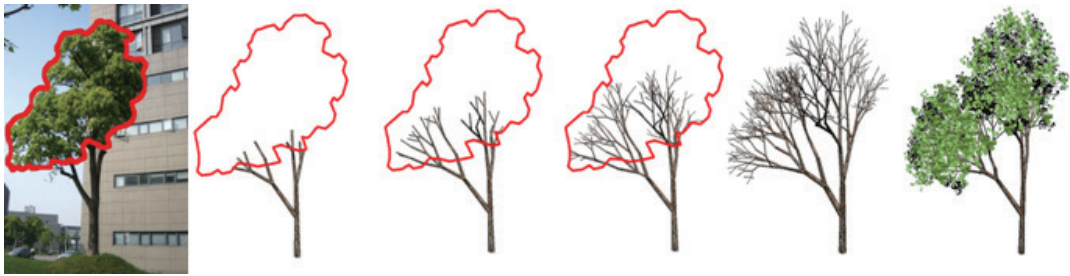


图 4-7 基于 L 系统 (L-Systems) 的 3D 植物生长建模 (图片来源：同济大学)



提示：我们通常喜欢设计有规则的图形。而分形 (Fractal)，则反其道而行之，研究的是复杂不规则的、支离破碎的形状，如图 4-8 所示，例如，弯弯曲曲的海岸线、起伏不平的山脉、变幻无常的浮云等。有趣的是，按照分形的观点，这些复杂对象虽然全局上看起来杂乱无章，但它们却具有**自相似性 (Self-Similarity)**：局部的形态放大后与整体的形态是相似的！以海岸线为例，在空中拍摄的 100km 海岸线与放大了 10 倍的 10km 海岸线的两张照片，看上去会十分相似！此外，分形的**分数维 (Fractional Dimension)**，通常为**豪斯多夫维数 Hausdorff Dimension**一般大于拓扑维数，用来度量形状复杂性和不规则性的程度：如下图最左边的科赫 (Koch) 雪花曲线，如果等比例放大 3 倍，周长变为 4 倍而不是 3 倍，则它的分数维 $D = \lg 4 / \lg 3 \approx 1.262$ ，大于线的拓扑维数 1 (另：面的拓扑维数为 2，体的拓扑维数为 3)，同时可注意到其不是一个整数。

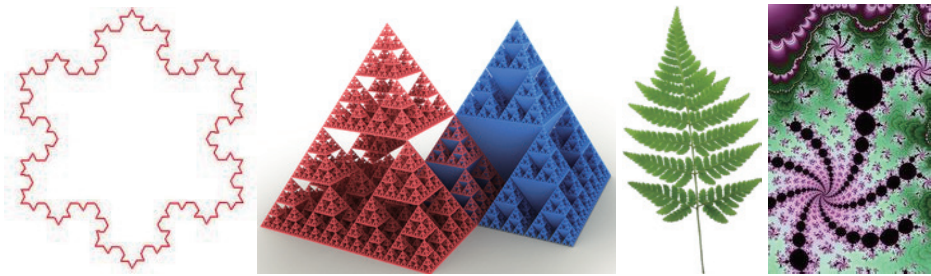


图 4-8 各式各样的分形图案 (图片来源：维基百科)

智能化达到一定层次后，更可让设计的形状根据未知环境实时调整，以适应各种物理或美学约束条件。因为，设计师是不可能提前知道最终的设计是什么样子的。例如，在崎岖不平的月球表面让 3D 打印机自动设计并制造房子，就需要根据所处的物理环境进行自适应性调整，以此来动态获得一个最优的设计形状，以保证在当前环境下建筑结构的稳定性。

采用人工智能进行设计的另一个途径是增强人和计算机之间的交互性，用户根本不需要了解计算机内部的运行原理，甚至不需要了解 3D 设计方法，只需从计算机推荐的参考形状中不断地做出挑选和评价 (“满意” 或 “不满意”)，然后计算机根据用户的反馈来分析用户的设计偏好，以此对参考形状进行优化调整，再重新推荐一个新的参考形状。如此反复，直到人和计算机共同合作完成一个满意的设计。在这个过程中，对人的设计水平要求大大降低，因为依托于计算机强大的分析能力和所存储的海量模型数据库 (含有上百万个 3D 模型)，用户的设计方式简单到只

需告诉计算机某个设计细节“满意”还是“不满意”即可,剩下的一切都交给智能的计算机算法了。

虽然 3D 数字化设计技术越来越先进,但目前唯一不变的还是大家手里握着的鼠标。鼠标是 2D 屏幕上的人机交互工具,面对 3D 空间的应用越来越力不从心。这里给大家介绍一款名叫“鸟标”的交互工具。在 CES 2013 展会上,出现了一款虚拟现实(Virtual Reality,简称 VR)系统:Leonar3Do VR suite。Leonar3Do 让用户能够在逼真的 3D 环境中设计形状。该技术的核心在于一款颠覆传统模式的 3D 鼠标:鸟标,它具有 6 个自由度(包括三维平移加三维转动),用于在 3D 空间中随心所欲地拖曳和移动 3D 模型,选取工具、颜色和纹理,挖除或雕刻塑像,添加或删除材料等,如图 4-9 所示。

为了达到精确逼真的视觉效果,鸟标需要配备头部运动跟踪装置和 3D 眼镜,以使计算机生成的 3D 场景跟使用者的视角相匹配。这样使用者就能像置身于一个真实的世界里那样编辑虚拟的 3D 物体,当编辑完成之后,可输出模型并将其 3D 打印成真正的物体。

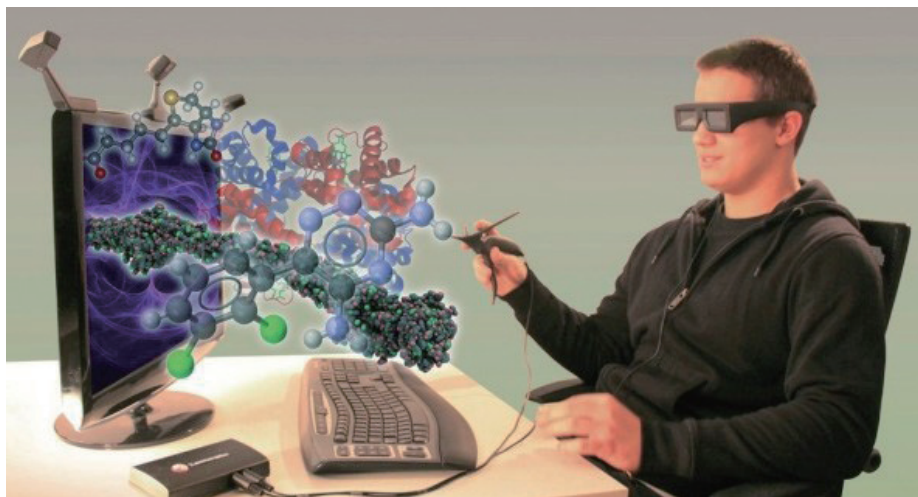


图 4-9 鸟标在 6 个自由度空间中自由操作,包括三维平移加三维转动(图片来源:Leonar3Do)

4.2.2 商业设计软件:3D 设计的重型武器(Maya、UG)

目前,各类商业化的 3D 设计软件在国内得到广泛应用,其大体可分为两大类:通用性 3D 设计软件和行业性 3D 设计软件。下面我们分别进行介绍,并做一些点评。

一、通用全能 3D 设计软件

3DS Max

3D Studio Max,常简称为 3DS Max 或 MAX,是当今世界上销售量最大的三维建模、动画及渲染软件。在 Windows NT 出现以前,工业级的 CG(Computer Graphics,计算机图形学)制作被 SGI 图形工作站所垄断。3D Studio Max + Windows NT 组合的出现一下子降低了 CG 制作的门槛,3DS Max 可以说是最容易上手的 3D 软件,其最开始应用于计算机游戏中的动画制作,后更进一步开始参与影视片的特效制作,例如,《X 战警 II》、《最后的武士》等。

Maya

Maya 应用对象是专业的影视广告、角色动画、电影特技等。Maya 功能完善，工作灵活，易学易用，制作效率高，渲染真实感强，是电影级别的高端制作软件，如图 4-10 所示。

Maya 集成了 Alias/Wavefront 先进的动画及数字效果技术。它不仅包括一般三维和视觉效果制作的功能，而且还与先进的建模、数字化布料模拟、毛发渲染、运动匹配技术相结合。



图 4-10 Maya 软件的界面 (图片来源: Autodesk)

Softimage

Softimage 是动画制作的顶级软件，和同类比起来最大的优点是输出质量好，原因是它集成了 Mental Ray 渲染器。Mental Ray 图像渲染软件由于有功能丰富的算法，图像质量优良，成为业界的主流。Softimage XSI 第一个将非线性概念引入到三维动画创作中。Softimage 曾经长时间垄断好莱坞电影特效的制作，在业界一直以其优秀的角色动画系统而闻名。

Rhino

Rhino (犀牛) 是美国 Robert McNeel 公司开发的专业 3D 造型软件，它可以广泛地应用于三维动画制作、工业制造、科学研究以及机械设计等领域。它能轻易整合 3DS Max 与 Softimage 的模型功能部分，特别是在创建 NURBS 曲线/曲面方面功能强大，对要求精细、弹性与复杂的 3D NURBS 模型，有点石成金的效能；能输出 OBJ、DXF、IGES、STL、3DM 等不同格式，并适用于几乎所有 3D 软件，尤其对增加整个 3D 工作团队的模型生产力有明显效果。

Blender

Blender 更符合创客开源精神，如图 4-11 所示，因为它是一款开源的跨平台全能三维动画制作软件，提供从建模、动画、材质、渲染到音频处理、视频剪辑等一系列动画短片制作解决方案；拥有方便在不同工作环境下使用的多种界面。Blender 以 Python 为内建脚本，支持 Yafaray 渲染器，

同时还内建游戏引擎，拥有极丰富的功能，强大的快捷键功能能让你事半功倍。

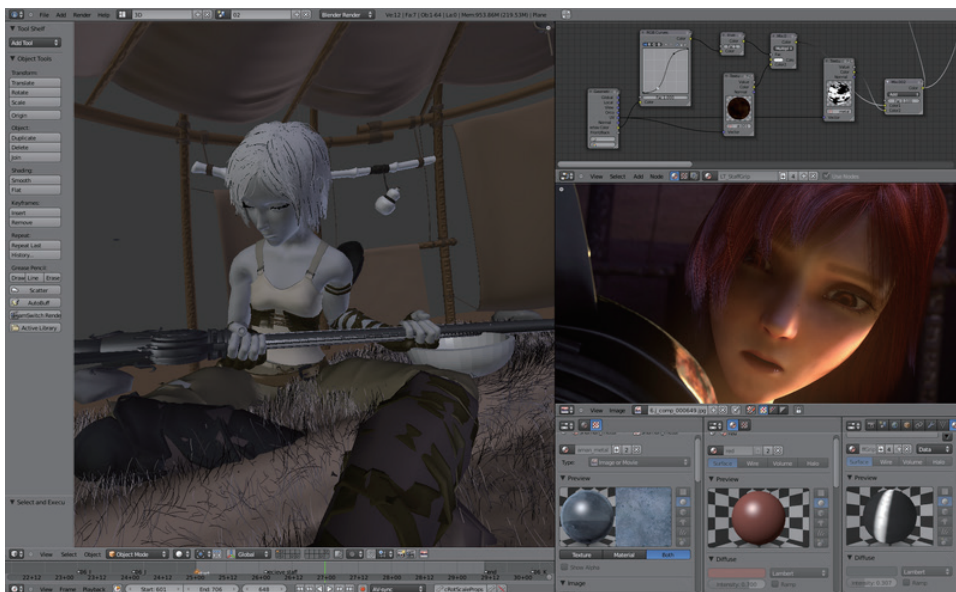


图 4-11 Blender 软件的界面

除了上面介绍的这些软件，更多的商业软件还有 LightWave 3D、Form-Z、Poser、ZBrush 等，在此不做详细介绍。其中有一些非常有特色的软件，比如 **Poser** 是一款专门针对三维人体 / 动物造型及动画制作的软件。又如 **Google SketchUp**（又名草图大师，分专业版和免费版），直接面向设计过程，以便建筑设计师直接和客户建立快捷的沟通，并能很方便地对构思设想做出及时修改。SketchUp 像使用铅笔一样简便，人人都可以快速上手，并且可以将创建的 3D 模型直接输出（如导出为扩展名为 dae 的 Collada 文件、扩展名为 kmz 的 Google Earth 文件）至 Google Earth 里。除此之外，**ZBrush/Autodesk MudBox/3D-Coat**（见第 5 章 5.5.1 节）是 3 款强大的 3D 雕塑软件，使得用户在进行 3D 建模时就像捏泥巴那样简单，可以随意地雕刻出极其复杂逼真的细节，比如层层叠叠的皱纹和褶子。

通用 3D 设计软件点评与比较：

不知该说是幸事还是憾事，最著名的 3 个软件（3DS Max、Maya、Softimage）现都被美国 Autodesk（欧特克）公司纳入囊中，因此 Autodesk 当之无愧地成为 3D 设计行业中的王者。

再看这几款软件，其实只要掌握任何一款基本上都能达到同一目的。如果你一定要“鸡蛋里挑骨头”，那我们可以点评一下。

Maya 和 3DS Max 比较如下。

- Maya 是高端 3D 软件；3DS Max 是中端软件，易学易用，但在遇到一些高级要求时（如角色动画、运动学模拟方面）远不如 Maya 强大。
- Maya 的用户界面也比 3DS Max 要人性化点。
- Maya 软件应用主要是动画片制作、电影制作、电视栏目包装、电视广告、游戏动画制作等。3DS Max 软件应用主要是动画片制作、游戏动画制作、建筑效果图、建筑动画等。Maya

的层次更高，3DS Max 属于普及型三维软件。

- Maya 的 CG 功能十分全面，包括建模、粒子系统、毛发生成、植物创建、衣料仿真等；可以说，当 3DS Max 用户匆忙地寻找第三方插件时，Maya 用户已经可以早早地安心工作了；可以说，从建模，到动画，到速度，Maya 都非常出色。Maya 主要是为了影视应用而研发的。

另一方面，Softimage 是面向高端三维影视市场的旗舰产品，以其独一无二、真正的非线性动画编辑功能为众多三维计算机艺术人员所喜爱。Softimage-XSI 最适合具有中到大型制作团队的公司级用户。如果团队中具有技术指导、建模师、动画师等合作工作的环境，Softimage-XSI 可以让所有项目的合作者在同一时间共同工作于一个场景当中。

二、行业性的 3D 设计软件

CATIA

CATIA 是法国达索 (Dassault System) 公司的 CAD/CAE/CAM 一体化软件。在 20 世纪 70 年代，CATIA 的第一个用户就是世界著名的航空航天企业 Dassault Aviation，后者以设计幻影 2000 和阵风战斗机最为著名。CATIA 源于航空航天业，但其强大的功能已得到各行业的认可，在欧洲汽车业，已成为事实上的标准。典型案例从大型的波音 747 飞机、火箭发动机到化妆品的包装盒，几乎涵盖了所有的制造业产品。CATIA 的著名用户包括波音、克莱斯勒、宝马、奔驰等一大批知名企业。波音飞机公司使用 CATIA 完成了整架波音 777 的电子装配，创造了业界的一个奇迹。

SolidWorks

SolidWorks 现也属于法国达索 (Dassault System) 公司，是面向中端主流市场的机械设计软件，如图 4-12 所示。SolidWorks 是基于 Windows 平台的全参数化特征造型软件，它可以十分方便地实现复杂的三维零件实体造型、复杂装配和生成工程图。图形界面友好，用户上手快。该软件可应用于以规则几何形体为主的机械产品设计，其价位适中。

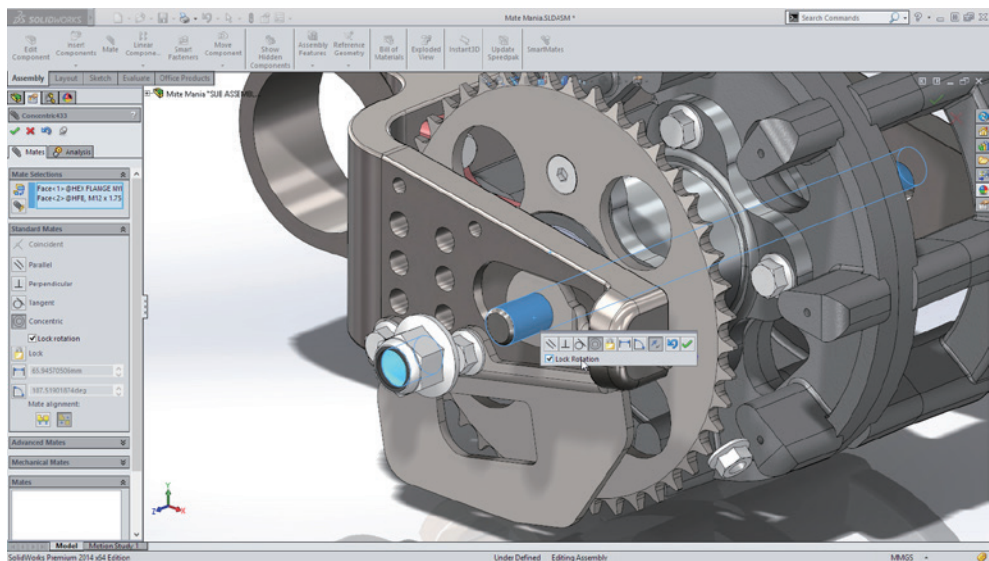


图 4-12 SolidWorks 软件的界面

Unigraphics (UG)

UG（现改名为 NX）目前属于 Siemens 公司。在 UG 中，优越的参数化和变量化技术与传统的实体、线框和表面功能结合在一起。UG 最早应用于美国麦道飞机公司。它是从二维绘图、数控加工编程、曲面造型等功能发展起来的软件。后来，美国通用汽车公司选中 UG 作为全公司的 CAD/CAE/CAM/CIM 主导系统，这进一步推动了 UG 的发展。

AutoCAD 与 MDT

AutoCAD 是 Autodesk 公司的主导产品。Autodesk 公司是世界第四大 PC 软件公司。目前在 CAD/CAE/CAM 工业领域内，该公司是全球规模最大的基于 PC 平台的 CAD 和动画及可视化软件企业。AutoCAD 是当今最流行的二维绘图软件，它在二维绘图领域拥有广泛的用户群。AutoCAD 具有强大的二维功能，如绘图、编辑、剖面线和图案绘制、尺寸标注以及二次开发等功能，同时也有部分三维功能。

AutoCAD 的强项在于二维绘图，而 MDT 是 Autodesk 公司在 PC 平台上开发的三维机械 CAD 系统。MDT 以三维设计为基础，集设计、分析、制造以及文档管理等多种功能为一体，为用户提供了从设计到制造一体化的解决方案。该软件的推出受到广大用户的普遍欢迎。由于 MDT 与 AutoCAD 同时出自 Autodesk 公司，因此两者完全融为一体，用户可以方便地实现三维向二维的转换。

Pro/Engineer

Pro/Engineer（简称 Pro/E，现改名为 Creo）是美国参数技术公司（Parametric Technology Corporation，简称 PTC）的产品。PTC 公司提出的参数化、基于特征、全相关的概念改变了机械 CAD/CAE/CAM 的传统观念，这种概念已成为了标准。利用该概念开发出来的 Pro/Engineer 软件能将设计至生产全过程集成到一起，让所有的用户能够同时进行同一产品的设计制造工作，即实现所谓的并行工程。PTC 近年又推出了 Creo，通过直接建模的全新概念来逐步取代 Pro/E 的参数化建模。

Cimatron

Cimatron 系统是以色列 Cimatron 公司的 CAD/CAE/CAM 产品。该系统提供了比较灵活的用户界面，优良的三维造型、工程绘图，全面的数控加工，各种通用、专用数据接口以及集成化的产品数据管理。Cimatron 系统在国际上的模具制造业备受欢迎。

CAXA 电子图板

CAXA 电子图板是我国国产 CAD 软件，如图 4-13 所示，由北京数码大方科技股份有限公司（原北京航空航天大学华正软件研究所）研发。该公司是从事 CAD/CAE/CAM 软件与工程服务的专业化公司。CAXA 电子图板是一套高效、方便、智能化的通用中文设计绘图软件，可帮助设计人员进行零件图、装配图、工艺图表、平面包装的设计，对我国的机械国家标准贯彻得比较全面。

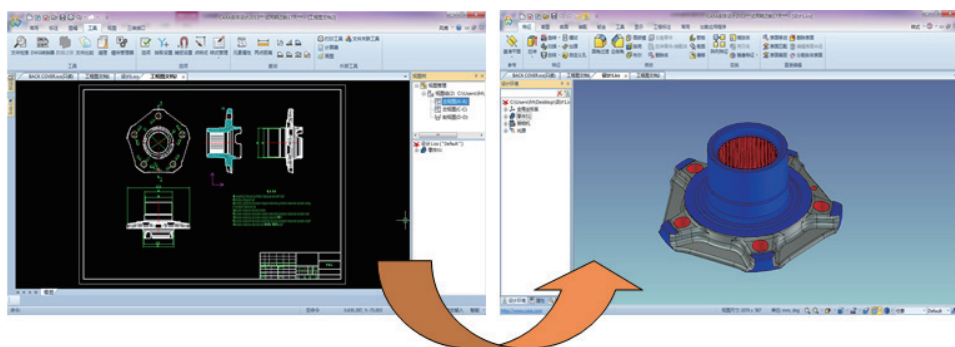


图 4-13 CAXA 软件的界面（由二维 CAD 图纸转换到三维零件）

开目 CAD

开目 CAD 是武汉开目信息技术有限责任公司开发的具有自主知识产权的 CAD 和图纸管理软件，它面向工程实际，操作简便，机械绘图效率高，符合我国设计人员的习惯。开目 CAD 支持多种几何约束种类及多视图同时驱动，具有局部参数化的功能，能够处理设计中的过约束和欠约束的情况。

除了上面的这些软件，更多的行业软件还有 I-DEAS、Solid Edge、Inventor、中望 3D、尧创 CAD、浩辰 CAD 等。

CAD 行业设计软件点评与比较：

一口气介绍了这么多行业 CAD 软件，可能大家都会有些茫然。行业设计软件现在确实有很多，目前用的最多的是 SolidWorks 软件，因为其上手快，设计和出图的速度快（采用 Windows 的技术，支持特征化的“剪切、复制、粘贴”操作），而且价格便宜。SolidWorks 兼容了中国国标，可以直接提取一些标准件和图框，不需要安装外挂。

Pro/E 属于中端三维软件，其界面简单，操作快速。Pro/E 的参数化建模在曲面建模时具有很大的曲线自由度，但同时也不容易很好地控制曲线。PRO/E 在装配设计方面也有长处，草图功能非 UG 所能比。

UG 属于高端三维软件，所以就功能来讲显而易见地强大，当然 UG 学习起来肯定要比 Pro/E 难一点。UG 可以集设计、加工、编程及分析于一体化，尤其在模具及加工编程方面比较突出。UG 的一个最大特点就是混合建模（参数化建模和非参数化建模混合操作）。

UG、CATIA 这两款软件在功能上比 Pro/E 要强得多，都普遍应用于汽车（奇瑞汽车用 CATIA，通用汽车用 UG）和航空领域，特别是 CATIA 作为波音公司的御用设计平台。这两款软件在曲面造型和 CAM 领域都有非常突出的优势，具备强大的曲线架构和编辑功能，在进行正向或逆向造型时得心应手。UG 的建模灵活，其混合建模功能强大，很多实体线条等都不必基于草图，中途可任意去参，无参修改相当方便，这些都是 CATIA 没法比的；UG Imageware 的逆向功能也是 CATIA 一下子难以超越的。

最后提一下 Cimatron。Cimatron 是伴随着大量的台湾企业进入大陆制鞋领域的。

当然，以上的这些软件具有很强的行业性，只有一些特殊行业的人才会用它们。

4.2.3 杀鸡焉用牛刀：基于网页的设计软件（Tinkercad、3DTin）

上一节我们介绍了商业化的3D专业设计软件，比如 Maya、SolidWorks、CATIA 等众多软件。这些软件功能强大，但同时学习门槛较高，特别是对完全没有设计基础的朋友来说学起来相当不容易，而且绝大部分是需要付费购买的，且价格不菲。因此，在本节中为大家介绍几款基于 WebGL 的轻量级 3D 设计工具，特点是无须安装，直接在 IE 网页浏览器中就可运行，而且是免费的。

Autodesk Tinkercad

Tinkercad 是一个基于 WebGL 的实体建模（Solid Modeling）网页应用，如图 4-14 所示，3D 建模功能非常简单，仅支持几种基本几何体（Primitive，如立方体、圆柱体等）以及基本几何体之间的布尔运算（如从一个立方体中间掏空一个圆柱体），如图 4-15 所示。

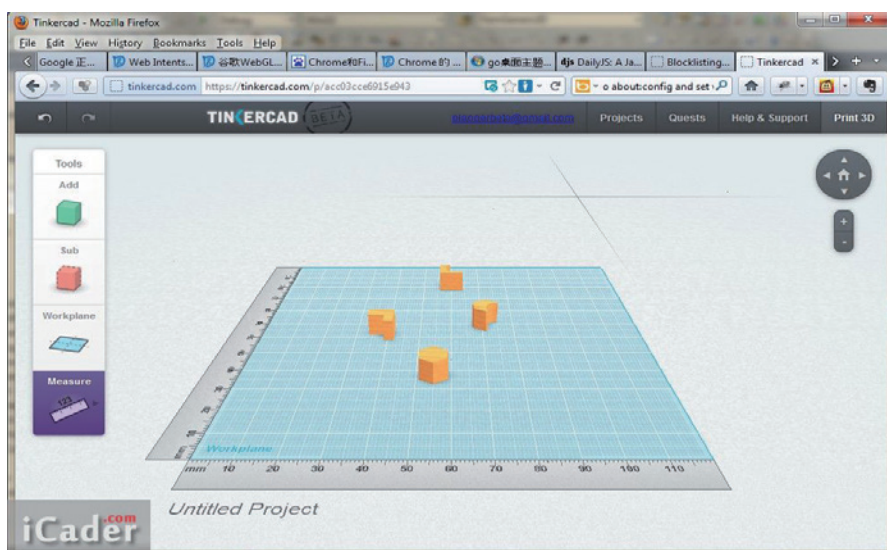


图 4-14 基于网页的 3D 设计软件 Tinkercad 的界面

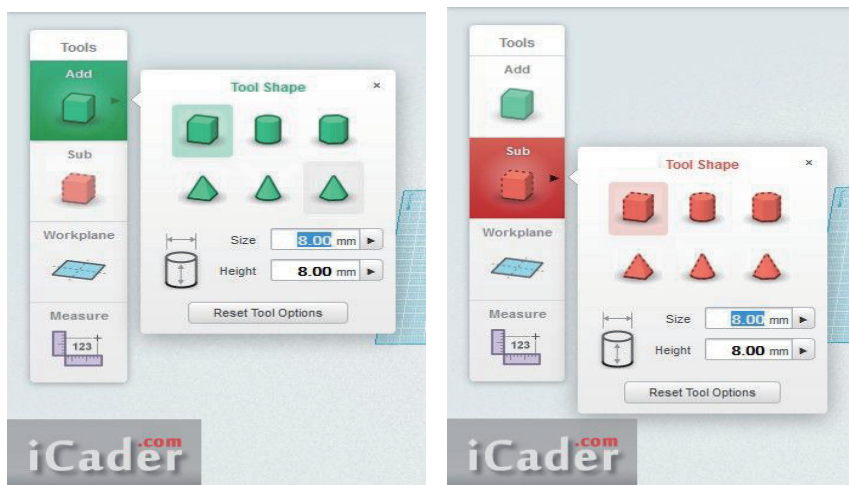


图 4-15 基本几何体的创建以及布尔运算（左：加法，右：减法）

寥寥数语之后,大家可能会对这款免费软件有些垂头丧气。不过别灰心,来看看榜样的力量!这里给大家展示一下由国外一名叫 Emily 的女创客使用 Tinkercad 建模然后打印出的酿酒屋,如图 4-16 所示。令人惊讶的是,她借用的是一张旧照片——1930 年的一张明信片上的酿酒屋照片——完成她的 3D 模型。当然,她也去实地获得了一些细节。



图 4-16 使用 Tinkercad 根据一张 1930 年的旧照片重建出的 3D 模型(图片来源:Emily)

成功的 3D 建模需要的是持久的耐力,特别是你用来“扬名立万、行走江湖”的建模作品。于是,Emily 使用 Tinkercad “一块砖接一块砖”地堆出了漂亮的细节和令人惊讶的外观表现,如图 4-17 所示。



图 4-17 将打印出来的模型与实物比较

通过这个案例说明,只要掌握了基本的方法,工具或装备已不那么重要,此时“无招胜有招、沾花飞叶皆可为剑”!当然,更多的是需要创造的热情,有将自己的设计或者想法变成实物的冲动,这样人人都可为创客!

3DTin

3DTin 也是一款使用 WebGL 技术开发的 3D 建模工具,你可以在浏览器中创建自己的 3D 模型。模型可以保存在云端或者导出为标准的 3D 文件格式,例如 Obj 或 Collada 格式。这里就不做详细介绍了,界面和效果如图 4-18 所示。

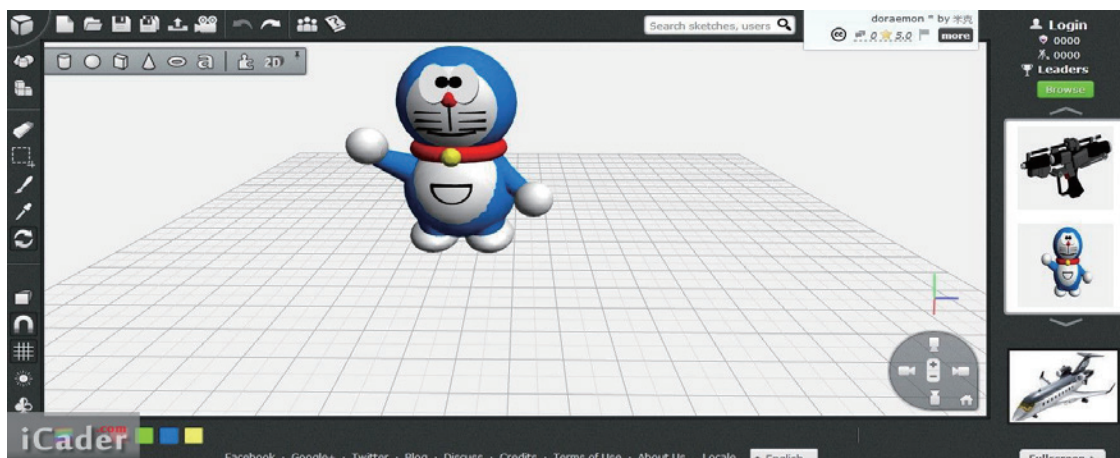


图 4-18 基于网页的 3D 设计软件 3DTin 界面

这两款基于网页的 3D 设计软件操作起来都特别简单，可谓傻瓜式、零基础直接上手。除了它们之外，还有一些免费的 3D 软件，如 Google SketchUp 免费版、Autodesk 123D、OpenSCAD、Art of Illusion、Sculptris、MakeHuman 等，操作会稍微复杂一些，同时还需要在计算机上安装。这些软件主要以功能的独特性来吸引用户，比如 Sculptris 是款免费的 3D 雕刻软件（是 ZBrush 的同门小师弟），小巧而实用，用户可以像玩橡皮泥一样，拉、捏、推、扭等来设计形状；再比如，MakeHuman 是一款专门针对人物制作、人体建模的 3D 软件，这款软件的亮点是可以让用户把玩身体和面部细节，保持肌肉运动的逼真度。

当然，如果你要设计的形状非常复杂，推荐还是使用 4.2.2 节中提到的商业化的 3D 设计专业软件。

4.3 3D 智能数字化扫描技术

在 4.2 节中我们详细介绍了 3D 智能数字化设计技术。然而，并非人人都有能力自己设计 3D 形状，因此第二大类的 3D 数字化就是 3D 扫描（俗称“3D 照相”、“3D 抄数”），利用计算机视觉、计算机图形学、模式识别与智能系统、光机电一体化控制等技术对现实存在的 3D 物体进行扫描采集，以获得逼真的数字化重建。

所谓 3D 扫描，说得通俗一点，就是类似于用我们日常的照相机对视野内的物体进行照相。区别在于照相机获得的是物体的二维图像，而 3D 扫描获得的是物体的三维信息。



注意：3D 扫描属于**逆向工程（Reverse Engineering）**的一种，通过扫描产品实物的 3D 外形来获取原本不公开的数字化设计图纸。这可能会涉及版权问题，比如将其他厂商生产的一款鞋子外形进行 3D 扫描（**3D 抄数**），逆向取得设计图纸后生产以获取商业利益。

研发人员通过扫描产品实物（如市面上已有的一款鼠标），通过逆向重建软件（如 Geomagic Studio）可迅速获得该款鼠标的 CAD 三维模型，并以此为基础，再借助正向 CAD 软件（如 SolidWorks）加入自己的创新设计元素，即可生产出一款新型的鼠标，这样大大节省了研发人员的设计时间、提高了工作效率。整个正逆向混合设计流程如图 4-19 所示。

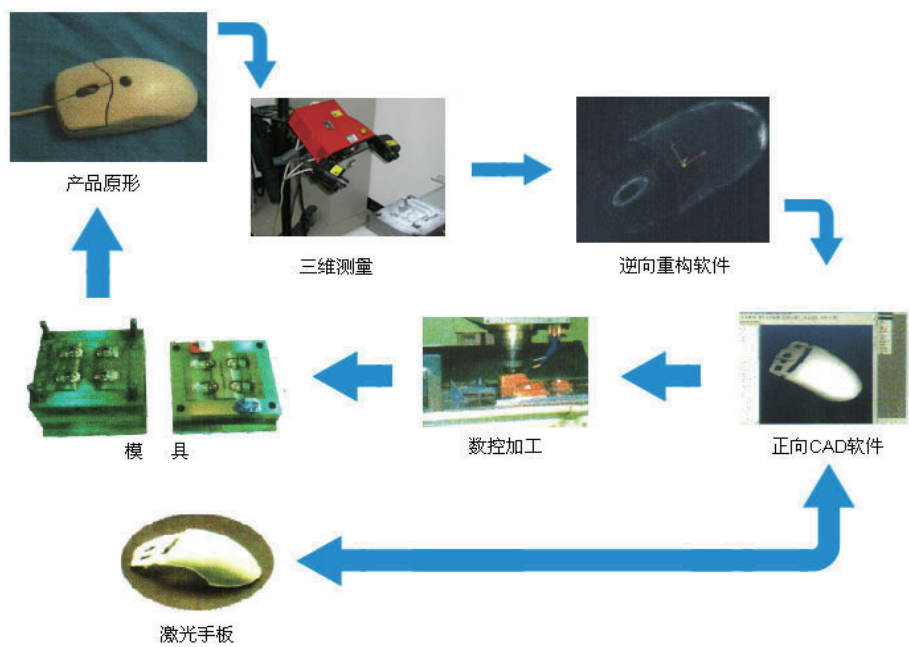


图 4-19 3D 扫描与 3D 设计相结合的产品研发流程（图片来源：深大三维）

3D 扫描及数字化系统可广泛应用于汽车、模具制造、家具、工业检测、制鞋、医疗手术、动漫娱乐、考古、文物保护、服装设计等行业，以提高行业生产效率。

雕塑行业的 3D 扫描应用（如图 4-20 所示）。



图 4-20 雕塑行业的 3D 扫描应用

家具行业的 3D 扫描应用（如图 4-21 所示）。

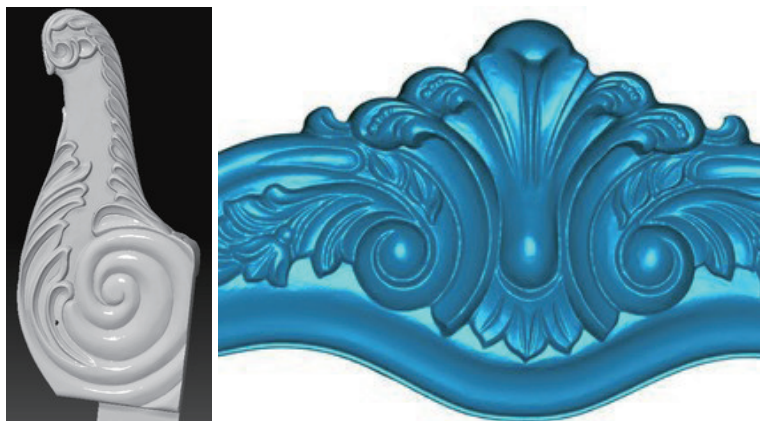


图 4-21 家具行业的 3D 扫描应用

大型零部件（发动机缸盖）逆向工程（如图 4-22 所示）。

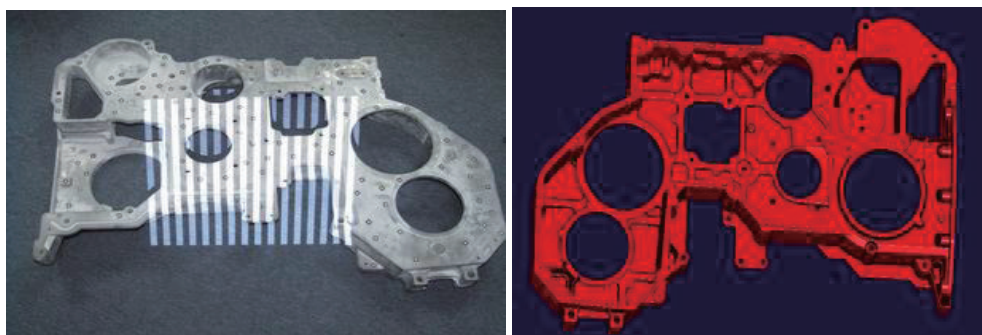


图 4-22 发动机缸盖的扫描现场和三维数字化模型（图片来源：华朗三维）

3D 扫描仪大体分为接触式和非接触式两大类。接触式三维扫描仪通过实际触碰物体表面的方式计算深度，如三坐标测量仪即典型的接触式三维测量仪。接触式测量仪精度很高，对物体表面的颜色、反射特性无要求；但缺点是，逐点测量速度慢且被测物有被探头破坏损毁的可能，因此不适用于高价值对象，如古文物、遗迹等。

非接触式三维扫描仪又分为结构光栅三维扫描仪（也称拍照式三维扫描仪）和激光扫描仪。光栅三维扫描又分白光扫描或蓝光扫描等，激光扫描仪又有点激光、线激光、面激光的区别。非接触式三维扫描仪由于采用逐线或逐面大范围扫描，所以扫描速度快而且精度高，但无法测量被遮挡的几何特征，因为要接收物体对光的反射，所以对零件表面的反光程度、颜色有要求（解决办法是可在表面喷涂白色显像剂），还容易受环境光线及散射光的影响。

这里对三维扫描仪的发展历程做一个简介。

第一代三维扫描仪：点测量

代表系统有：固定式三坐标测量机、便携式关节臂测量机、点激光测量仪。通过每一次的测量点来求得物体的表面特征，优点是精度特别高，但速度慢。

适用范围：适合做物体表面误差检测用。

第二代三维扫描仪：线扫描

代表系统有：台式三维激光扫描仪、手持式三维激光扫描仪、关节臂测量机配激光扫描测头。通过一段（不能过长，否则激光线会发散）有效的激光线照射物体表面，再通过传感器得到物体表面的三维信息。

适用范围：适合扫描中小件物体，扫描景深小，精度较低，这一代系统发展已比较成熟，但受原理的局限，目前已属于过渡性产品。

第三代三维扫描仪：面扫描

代表系统：“拍照式”结构光三维扫描仪，三维摄影测量系统等。通过一组（一面光）光栅的位移，再同时经过传感器而采集到物体表面的数据信息。

适用范围：应用范围非常广泛，相比于激光的线扫描，面扫描速度更快、精度更高。

4.3.1 光学三维扫描仪的原理和实例（激光、结构白光）

目前光学三维扫描仪按照其原理分为两类，一种是“拍照式”，一种是“激光式”，两者都是非接触式的，采用的是三角测距原理。以激光为光源，将激光束投射到物体表面，用 CCD 数码相机在另一位置接收激光的反射。这样，光源、物体、CCD 数码相机三者之间就形成了一个三角形。因此，我们可以利用光源和 CCD 数码相机之间的位置和角度关系来计算物体表面点的三维坐标。激光三维扫描可以分为点扫描法和线扫描法两种，如图 4-23 所示。

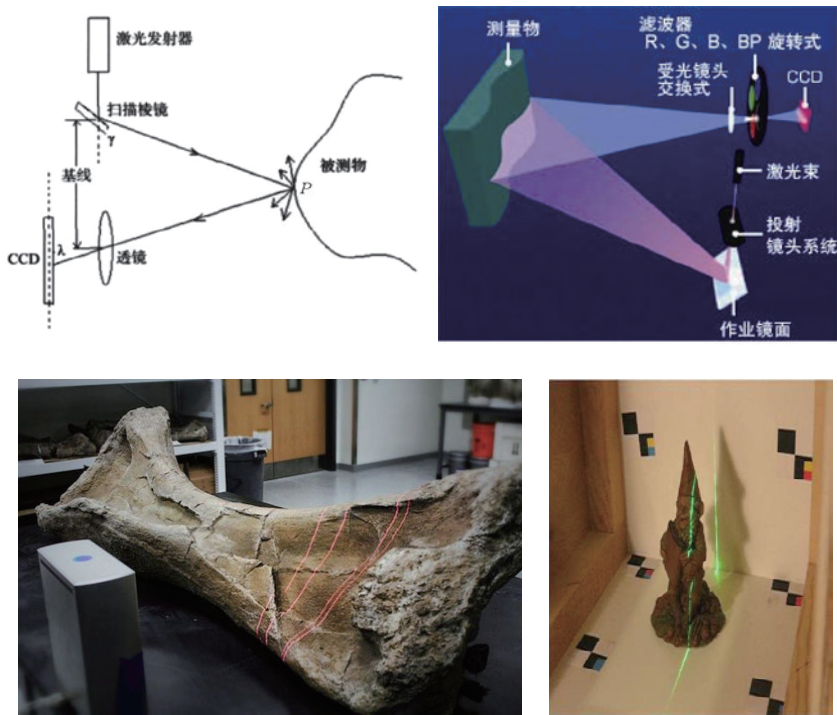


图 4-23 激光三维扫描原理。左上：点扫描法；右上：线扫描法；下：两个线扫描示例（图片来源：instructables.com）

“拍照式”结构光扫描仪是针对工业产品设计领域的新一代扫描仪，如图 4-24 所示，与传统的激光扫描仪的线扫描比较，面扫描的测量速度提高了很多，整体测量精度也有提高，扫描范围可达 10m。两天时间即可完成整车的内外表面扫描。

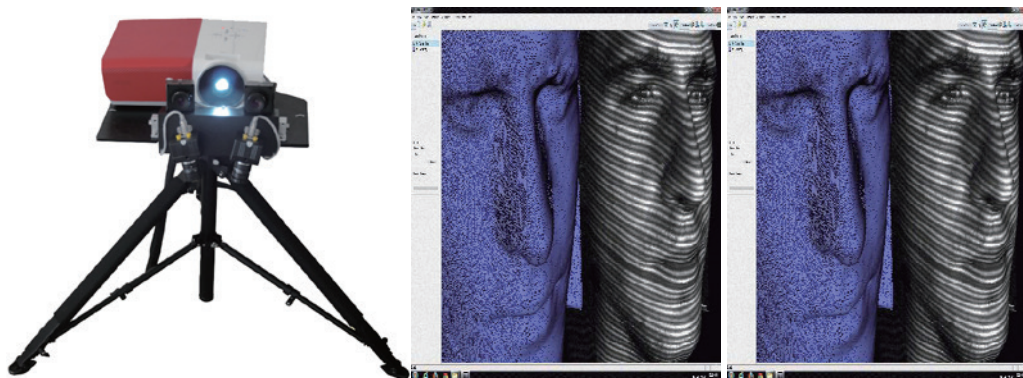


图 4-24 “拍照式”结构光扫描仪（图片来源：Alan Robinson）

结构光三维扫描仪的主要结构，由光栅投影设备及两个成一定夹角的高分辨率 CCD 数码相机所组成，如图 4-25 所示。工作原理是：将编码后的数字光栅条纹（如采用正弦明暗分布）投影在被测物表面上，光栅条纹受到物体表面高度的影响而发生变形（条纹间的正弦相位关系发生了变化），变形后的光栅条纹因此携带了物体表面的三维信息。然后由两个 CCD 数码相机获取不同角度的数字图像，经计算机算法（利用立体匹配技术、三角形测量原理、相位计算等）解码出光栅条纹变形所引起的相位变化量，即可重建出物体的 3D 形状。

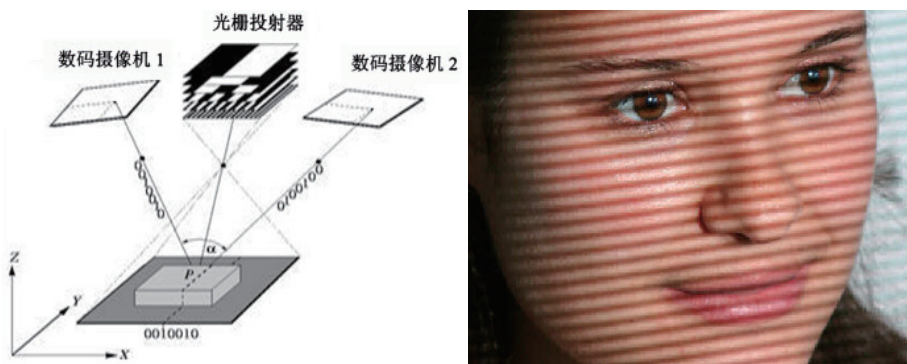


图 4-25 结构光三维扫描仪原理示意图（图片来源：polymetric.de）



提示：用两个 CCD 数码摄像机的好处是：物体、两个 CCD 数码摄像机三者之间直接进行三角测距。当然也可以只用一个 CCD 数码摄像机，但缺点是此时要根据物体、投影机、CCD 数码摄像机这三者进行三角测距，然而投影机一般是不便于固定死的，若其位置有错位就会导致系统误差。

每次扫描得到物体一个侧面的形状，于是将物体转动一个角度后再扫描新的侧面，直到得到物体 360° 各个方位的形状。两次扫描的侧面之间要保证一小部分形状是重叠的，以便于将两个侧面拼接起来。为了便于拼接，有时还需要在物体上随机（一定不能有规律，如有规律地连

成一条线) 贴上一些标记点 (Markers), 如图 4-26 所示, 邻接两个扫描面的公共标记点一般最少为 3 个, 常见的为 4 个或以上。由于标记点遮盖住的形状将变成空洞 (其只能通过后期修复), 因此标记点一般贴在物体的平坦无特征处。



图 4-26 为了便于将各个侧面拼接起来, 在物体上贴上标记点 (图片来源: 合肥智泰)



注意: 光学三维扫描适用于测量表面相对平坦的物体。在陡变不连续的曲面以及窄缝、边界处, 可能会发生相位突变而造成细节丢失。

下面这张图片给出了结构光三维扫描的工作流程。首先把物体放在转盘上, 并对其投射事先设计好的结构光条纹。为了唯一确定空间点的三维位置, 需要投射有一定相位差的多幅光栅条纹, 每幅条纹的粗细和位置都有不同, 参见图 4-27 上方的中间和右边图片。当扫描完一个侧面, 将转盘转一个角度再扫描下一个侧面, 要保证两个侧面的形状有部分重叠; 然后继续旋转、扫描, 我们就获得了多个侧面的三维形状数据。最后将这些侧面拼接 (也称为配准) 在一个统一的世界坐标系下 (具体步骤参考第 5 章 5.5.3 节), 于是我们就得到了一个完整的 360° 全方位 3D 模型。

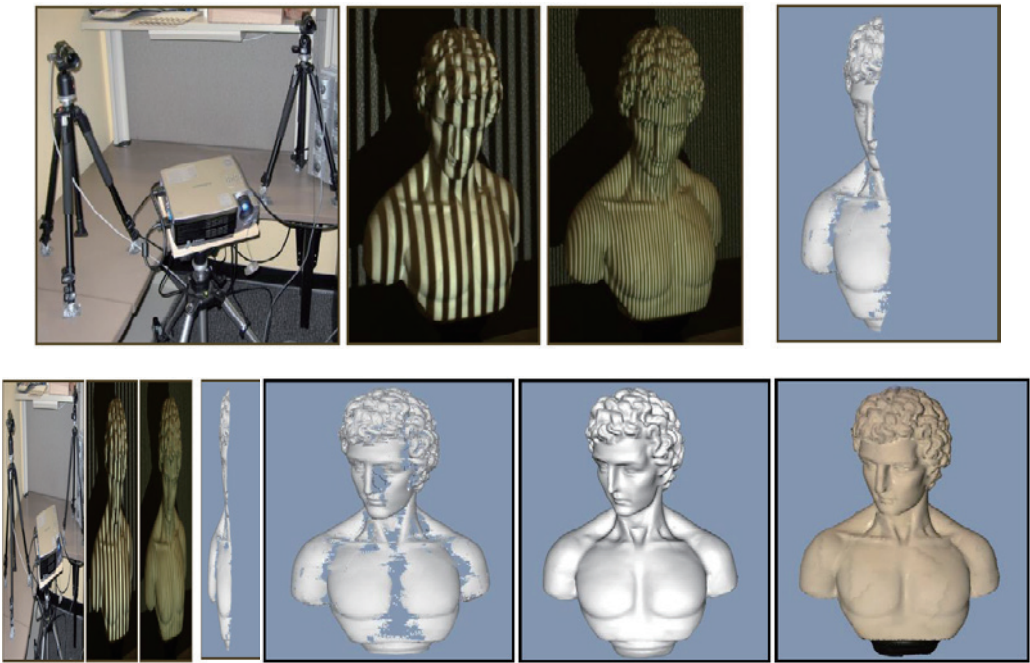


图 4-27 结构光三维扫描的工作流程 (图片来源: gdiy.com)

扫描后的形状拼接、形状优化、NURBS 曲面生成等过程需要配备专用的逆向软件。目前商业化的逆向工程软件有：Geomagic、Imageware、Polyworks、CopyCAD、ICEMSurf、RE-Soft、Rapidform 等。

从原理上讲，结构光测量法一般基于相位偏移（Phase Shifting）测量轮廓术（Measuring Profilometry），如四步相位移算法。这种采用黑白明暗分布的正弦条纹的优点是简单方便，但缺点是需要连续拍摄多张粗细及位移变化（即有一定相位移）的变形条纹照片才能计算出三维信息。



提示：以标准的四步相位移算法为例，4 幅光栅图像的相位移分别为 0 、 $\pi/2$ 、 π 和 $3\pi/2$ ，则 4 幅图像的光强分别为：

$$\begin{aligned} I_1(x, y) &= I'(x, y) + I''(x, y) \cos[\phi(x, y)] \\ I_2(x, y) &= I'(x, y) + I''(x, y) \cos[\phi(x, y) + \pi/2] \\ I_3(x, y) &= I'(x, y) + I''(x, y) \cos[\phi(x, y) + \pi] \\ I_4(x, y) &= I'(x, y) + I''(x, y) \cos[\phi(x, y) + 3\pi/2] \end{aligned}$$

其中的 3 个未知量： $I'(x, y)$ 为图像的平均灰度、 $I''(x, y)$ 为图像的灰度调制、 $\phi(x, y)$ 为待求解的**相对相位值**（也被称为**相位主值**）。

显然，相对相位值 $\phi(x, y)$ 可由以上的 4 个式子直接求解得出：

$$\phi(x, y) = \arctan\left(\frac{I_4 - I_2}{I_1 - I_3}\right)$$

然而，虽然相对相位值 $\phi(x, y)$ 在一个相位周期内唯一，但在多个光栅条纹的情况下呈锯齿起伏分布，因此还需进一步进行**相位展开（Phase Unwrapping）**以得到连续的**绝对相位值 $\Phi(x, y)$** ，比如采用时间相位展开算法。最后，我们就可以根据预先标定的系统参数或相位 - 高度映射关系，从绝对相位值计算出被测物体表面的三维点云坐标了。

举个例子，假设我们连续投射 6 组粗细不同的条纹，而每组都位移变化了 4 次，这样每个 CCD 数码摄像机需拍摄 $6 \times 4 = 24$ 张条纹图像，两个 CCD 相机共需拍 48 张图像，这么多张图像的处理时间需要长达 1s，因此黑白条纹法一般只适合静止不动的物件。如果要实时（最多 $1/25 = 0.04s$ ）重建出运动物体的 3D 形状，比如将脸被人扇一巴掌的形变过程扫描下来（如图 4-28 所示），则需要原理上的改进。

一种思路是将黑白条纹更改成彩色条纹，如图 4-29 所示。编码方式也做一些调整，如改为 De Bruijn 序列。这时只需拍摄一张（One-shot）彩色条纹图案的照片即可重建出 3D 信息，使实时扫描运动物体的 3D 形状序列成为了可能。



图 4-28 运动形变物体的实时扫描(图片来源:ETH Zurich)

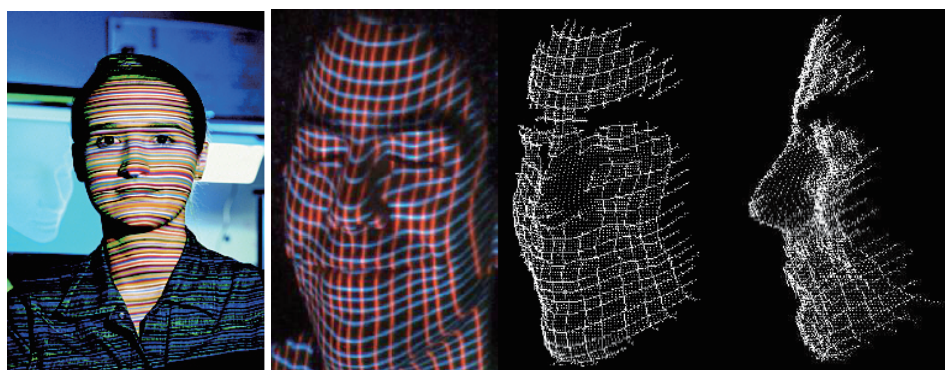


图 4-29 彩色条纹光以及实时获取的 3D 动态人脸表情(“笑”)(图片来源:Brown University)

最后,我们介绍美国的一款名叫 Artec 的结构光三维扫描仪,目前被 3D 照相馆广泛采用,但缺点是价格昂贵。如图 4-30 所示的是 Adidas Originals 工作团队为香港歌星陈奕迅(Eason)所做的人体 3D 扫描过程,由于扫描需要一定时间,请注意在手臂的着力点处架设了支架(在数据后处理中可去掉这个辅助支架),以避免长时间保持一个姿势的肌肉酸痛。此外在这个扫描案例中,两名工作人员一前一后同时进行扫描,这样进一步缩短了扫描的整体时间。





图 4-30 手持式结构光三维扫描仪的扫描案例（图片来源：Adidas）

4.3.2 基于 Kinect 的 3D 扫描原理和设备（红外光斑、ToF）

在 4.3.1 节我们介绍了两种光学扫描仪：激光和结构光。这两种 3D 扫描仪精度高，但有一些共同的缺点，比如高强度激光对人的眼睛有伤害，而高强度结构白光虽然无害，但也非常地刺眼。此外，它们的扫描速度也不是特别快，单个面的扫描时间一般都是秒级的，还没有快到毫秒级的。

因此在本节中，我们介绍一种结合激光和结构光优点的新式扫描仪：红外激光扫描仪，优点是对人眼无害且速度非常快，当然，缺点是精度要差一些，但对于 3D 照相馆这样的非工业级应用，目前基本可以胜任了。

红外激光扫描仪大体可以划分为两类。

基于飞行时间（ToF）原理

ToF 通过测量光脉冲之间的传输延迟时间来计算深度信息。ToF 是“Time of Flight”的缩写，从字面上也可知道其含义：计算光线飞行的时间。首先让装置发出脉冲光，同时在发射处接收目标物的反射光，即可通过测量时间差来算出目标物的距离。具体公式为：

$$\text{目标物的距离} = \text{光的往返飞行时间} \times \text{光速} / 2$$

感光芯片需要飞秒级（1 飞秒只有 1s 的一千万亿分之一，即 $1\text{e}-15$ 秒）的快门来测量光的飞行时间。ToF 的代表性产品有 Mesa Imaging SwissRanger 4000、PMD Technologies CamCube 2.0 等。

基于结构光编码原理

在 4.3.1 节我们已经介绍过，“结构光”指一些具有特定模式的光，其模式图案可以是点、线、面等。结构光扫描法的原理是首先将结构光投射至物体表面，再使用摄像机接收该物体表面反射的结构光图案。由于接收到的图案会因物体的立体形状而发生变形，所以可通过该图案的所处位

置和变形程度来计算物体表面的空间信息。红外结构光编码的代表性产品有以色列 PrimeSense 公司的 PrimeSensor 和 Microsoft Kinect (实际上, Kinect 就是微软向 PrimeSense 公司购买的同一技术)。

在本节中,我们重点介绍一下目前市面上广泛使用且价格低廉的 Microsoft Kinect。Kinect 是微软在 2010 年 6 月 14 日对 Xbox 360 体感周边外设正式发布的名字。Kinect 将人机自然交互带入全新领域,无论是游戏或应用,都可以通过身体姿态来控制。在 3D 打印出现后,技术人员发现了 Kinect 另一个功用:可以作为 3D 扫描仪,而且还可实现实时、动态地扫描。

Kinect 获得深度数据的原理如下:红外投影机的普通激光源投射出一道“一类普通激光”(Class1 Laser),这道激光经过磨砂玻璃和红外滤光片,覆盖 Kinect 的可视范围,然后红外摄像头接收反射光线,以此来获得目标物体的“深度场”(Depth Field),如图 4-31 所示。



图 4-31 Kinect 获得的 3D 深度场数据,不同颜色代表着不同的深度距离(图中为笔者本人)

上面这句话对于普通读者来说,可能说了等于白说。因为包含了太多的术语且不加以解释。下面,我们就抽丝剥茧般地进行详细解释。

在大学物理实验课上,我们都做过散斑实验。激光照射到粗糙物体或是穿透毛玻璃后,会形成随机的反射斑点,称为激光散斑(Laser Speckles)。这些散斑会随着距离的不同变换图案:空间中任意两处散斑图案都是不同的。

激光散斑因为是随机分布的,要研究它必须使用概率统计的方法。通过统计研究,可以认识到散斑的强度分布和运动规律。最核心的本质还是刚才那句话:散斑具有高度的随机性,会随着距离的不同出现不同的图案,且空间任意两点的散斑都不相同。因此,只要在空间中打上这样的散斑,就相当于给整个空间的远近位置做了标记。这样,我们从物体的散斑图案变化就可以得知该物体所处的空间位置。

Kinect 采用的光斑编码(Light Coding)就基于这个原理,如图 4-32 所示。可以看出,它与上节提到的周期性变化的二维结构光栅编码有所不同:光斑编码是直接具有三维纵深的“体编码”,只要看物体表面的光斑图案,就可以知道这个物体在什么位置。



图 4-32 通过红外夜视摄像机观察 Kinect 传感器投射出来的光斑（图片来源：zhihu.com）

当然，在这之前要把整个空间的散斑图案都记录下来，所以要先做一次光源的标定。方法如下：每隔一段距离，取一个参考平面，把参考平面上的散斑图案记录下来。假设 Kinect 规定的用户活动范围是距离电视机 1 ~ 4m，每隔 10cm 取一个参考平面，标定后保存了 30 幅散斑图像。这些标定结果在 Kinect 出厂前就已经固化在芯片中了。

用户购买 Kinect 后，实际测量时，Kinect 会拍摄一幅待测物体的散斑图像，将这幅图像与出厂前就已标定了的 30 幅参考图像依次做**互相关（Cross-Correlation）**运算，得到 30 幅相关度图像。在每一幅相关度图像上就会有相关性最大的峰值，这些峰值代表此处有物体存在的可能性最大，比如你的手出现在第 1 幅图像上的某个峰值坐标位置，你的脚出现在第 5 幅图像上的另一个峰值坐标位置等。把这些峰值一层层叠在一起，经过插值运算，并参照标定散斑图案时所测量的位置距离（比如第 1 幅图像的位置在 1m 处，第 5 幅图像的位置在 1.4m 处），即可得到整个物体的 3D 深度距离。整个原理是不是特别简单？



提示：**互相关函数（Cross-Correlation）**表示不同过程的某一时刻的相互依赖关系或匹配程度，而**自相关函数（Auto-Correlation）**则表示了同一过程不同时刻的相互依赖关系。设两个函数分别为 $f(t)$ 和 $g(t)$ ，则**互相关函数**定义为：

$$R(u) = f(t) * g(-t),$$

其中，*代表**卷积（Convolution）**。卷积可被理解为“加权平均积”，即一个函数对另

一个函数做加权平均： $c_{xy}(n) = x(i) * y(i) = \sum_{i=-\infty}^{+\infty} x(i)y(n-i)$ 。卷积的一个重要性质是：

时间域 / 空间域上的卷积对应于频率域上的乘积，即两个函数卷积后的傅里叶变换等于它们的傅里叶变换的乘积，这可使傅里叶分析中许多问题的运算得到简化。

自相关函数的定义则显得很简单：设原函数是 $f(t)$ ，则 $R(u) = f(t) * f(-t)$ 。

看到“激光”，你可能还会心存疑虑：Kinect 发射的是红外线还是激光？会不会对人体有害呀？实际上，Kinect 发射的是近红外激光，且符合 IEC 60825-1 标准中的一级（Class1）安全要求，大家大可放心。



参考：激光的波长范围包括：长波长的远红外激光（CO2 激光器），近红外激光（Nd:Yag、Nd:YVO4 激光器），可见光（氦、氩激光器）和不可见的紫外光。在光谱中波长 0.76 ~ 400mm 的一段称为红外线，红外线是不可见光线。所有的物质都可以产生红外线，其中有体温的生命体（如人类）发出的光谱属于“远红外线”，而 Kinect 发射的光谱属于“近红外线”。

根据对人的伤害程度可以定义激光不同的危险等级：从激光一级（在各种情况下，一级激光基本安全）到激光四级（任何情况下，四级激光是危险的）。

介绍了这么长时间，可能有些读者还没有见过 Kinect 的真容。下面就对 Kinect 关键部件做一个说明，如图 4-33 所示。



图 4-33 Kinect 的整体外形结构

将 Kinect 大卸八块后的内部结构如图 4-34 所示。



图 4-34 Kinect 拆解后的内部结构（图片来源：iFixit）

Kinect 以每秒 30 帧的高速生成 3D 深度图像，也即延迟仅为 $1/30\text{s}$ (33ms)，因此可以实时地再现周围 3D 环境和捕捉动态运动目标。Kinect 共有 3 个摄像头，中间的镜头是 RGB 彩色摄像头，左右两边的镜头分别为红外线发射器和红外线 CMOS 深度摄像头。深度感应器的有效视野范围是 0.8 ~ 3.5m。Kinect 在距离 1m 时精度大概是 3mm；而当距离是 3m 时，精度大概是 3cm，因此随着距离的拉远而精度降低。

具体来讲，Kinect 由以下部件组成。

- 红外线发射器：主动投射近红外光谱，照射到粗糙物体或是穿透毛玻璃后，光谱发生扭曲，会形成随机的反射斑点（称为散斑），进而能被红外摄像头读取。
- 红外深度摄像头：分析红外光谱，创建可视范围内的人体、物体的深度图像。分辨率为 640×480 。



提示：相比于 Kinect for Xbox 360，新版 Kinect for Windows 固件做了升级以改善 USB 传输瓶颈，深度分辨率变成 640×480 ，而 Xbox 360 的深度分辨率仅为 320×240 。此外，Kinect for Windows 还支持“近景模式”，可视范围为 0.4 ~ 3.5m。

- 仰角控制马达：Kinect 搭配了追焦技术，底座马达会随着对焦物体的移动跟着转动，用于获取最佳视角。
- RGB 彩色摄像头：用于拍摄视角范围内的彩色视频图像。分辨率为 640×480 。
- 麦克风阵列：声音从 4 个麦克风采集，同时过滤背景噪声，可定位声源。
- USB 线缆：支持 USB 2.0 接口，用于传输彩色视频流、深度流、音频流等，必须使用外部电源，传感器才能充分发挥其功能。（Kinect 的功率达到了 12W，而普通 USB 一般是 2.5W。）

Kinect for Xbox 360 不能离扫描对象太近，否则采集不到数据，形象点说，此时镜头对不上焦。而有时我们恰恰需要在空间狭窄的地方采集。针对“空间狭窄”这个问题，Nyko 公司推出可以“变焦”的 Kinect 配件“Zoom of Kinect”（Kinect 放大镜）。它基于鱼镜头的设计原理，将感应距离拉近并且向左右伸展，且没有影响到 Kinect 的感应精确度，如图 4-35 所示。这也可以理解为另外一种“近景模式”的解决方案，你可以在网上买到该配件。



图 4-35 Kinect 配件“Zoom of Kinect”（图片来源：Nyko）

下面介绍一下 Kinect 的 3D 扫描操作过程。Kinect 体积很小，可以方便地进行手持式扫描。当然，你也可以将 Kinect 放在一个能上下滑动的垂直支架上（便于从上到下扫描人体），并将扫描对象放置在可 360° 旋转的圆形电动转台上，如图 4-36 所示。



图 4-36 将 Kinect 放置在支架上扫描人体（图片来源：mbot3d）

值得指出的是，目前第一代的 Kinect 扫描精度不是很高（一般不高于 1mm），输出的原始 3D 数据有很多噪声，如果直接用于人像扫描则效果欠佳。这时，轮到 3D 智能数字化技术大显身手了。可结合数字滤波，加上 3D 模型库数据驱动的方法，使原始的粗糙表面“旧貌换新颜”，如图 4-37 所示。

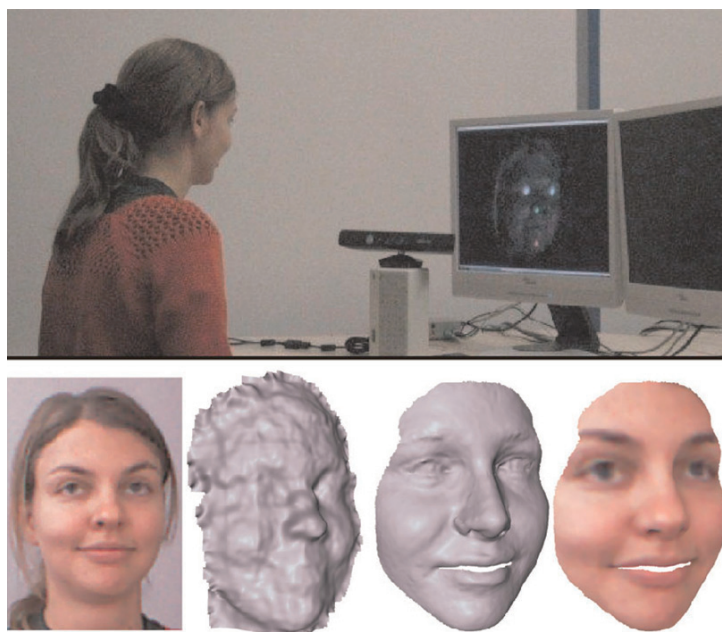


图 4-37 3D 智能数字化技术修复原始粗糙表面（下图第 2 个），获得细致表面（下图第 3 个）
（图片来源：University Erlangen-Nuremberg）

目前为 Kinect 配备的 3D 扫描软件有微软自带的 Kinect Fusion、Skanect（详见第 5 章 5.4 节“Skanect：使用 Kinect 实现 3D 扫描”）、ReconstructMe、Artec Studio 等。

最后简单介绍一下微软新设计的 Kinect 2。令人不感意外的是，微软作为一个大公司，在新一代 Kinect 的研发上，不再购买 PrimeSense 公司的技术，而是选择了自行研发。因此 Kinect 2 的原理与第一代 Kinect 是不同的，改为使用 ToF 代替结构光斑，精度也变为上一代的 3 倍，以至于可以区分用户的手掌和拇指。此外，RGB 彩色摄像头也提升到 1080p 高清模式。Kinect 2 的产品外形如图 4-38 所示。



图 4-38 Kinect 2 外形图（图片来源：tgbus）

4.3.3 房地产行业的新应用：室内 3D 扫描建模

在本节中，我们将 3D 扫描更拉近我们的日常生活：房屋家居。美国硅谷初创公司 Matterport 发明了一种新的小型扫描仪。这款小型扫描仪能够瞬间扫描用户所在的环境空间（比如你的房间）并建立 3D 模型，且建立模型的过程是实时的。

这款小型扫描仪安装在一个三脚架上，头部包含了两个摄像头。它使用 PrimeSense 感应器来自动产生美观、彩色的室内空间的 3D 模型，和微软的 Kinect 采用相同的技术。当你按下扫描仪头部的“扫描”按钮，然后原地转 360°，扫描就完成了，如图 4-39 所示。用户可以通过 iPad 来查看和控制扫描得到的 3D 模型。

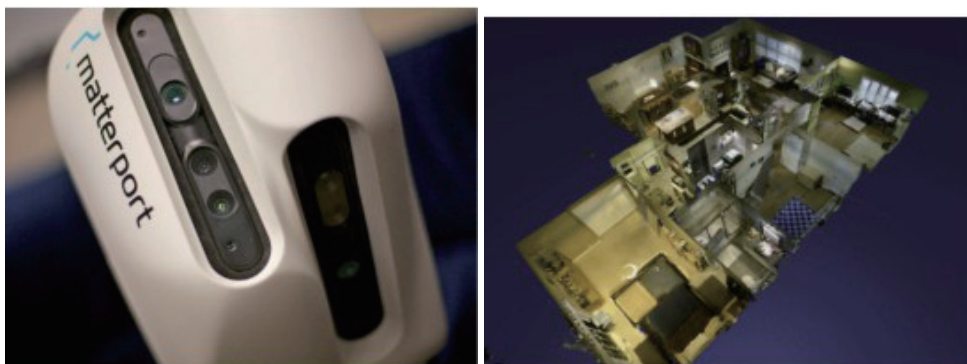


图 4-39 Matterport 公司的房间扫描仪以及扫描案例

“前人栽树，后人乘凉”。基于 Matterport 的扫描仪，纽约一家名为 Floored 的公司跟着沾光了，一下募集到 100 万美元，为房地产业开发了专业的 3D 建模软件，如图 4-40 所示。该软件能让用户自由地修改模型，例如，挪动家具、墙壁，增加窗户使室内获得更多阳光等。这样一来，用户就能在屏幕上按自己的想象，随意调整室内装修。

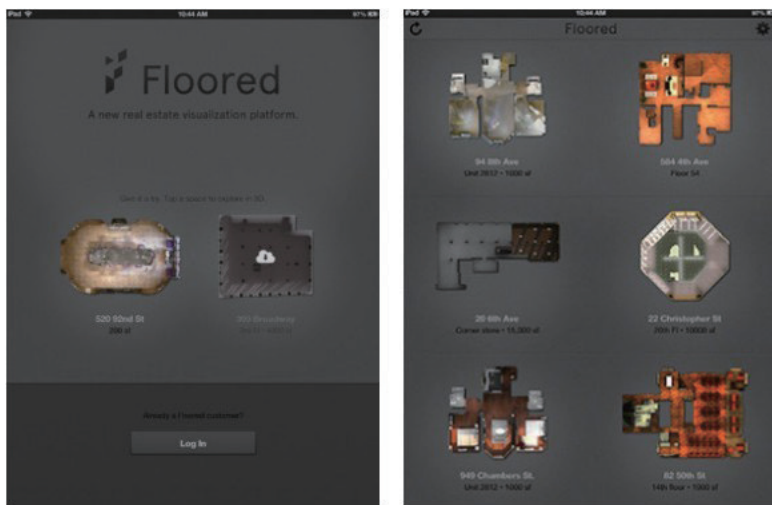


图 4-40 Floored 募集到 100 万美元为房地产业开发 3D 建模软件

Floored 公司的创始人 Eisenberg 认为：“3D 数字扫描和 3D 打印（的发展）是平行的。一旦 3D 打印机进入到主流市场，并成为普通的家用器具之后，3D 打印的价值就将从其周边软件的功能，从购买、创造和共享设计的方式等这些方面得到进一步体现。Floored 认为 3D 扫描领域也将经历同样的局面。”

4.4 面向“批量定制”和“柔性制造”的智能数字化

未来的制造会是什么样的？也许目前大规模“**批量生产**”（Mass Production，大规模生产）的工业产品将越来越少，替代的是个性化定制的创意设计产品。这是伴随着互联网、智能数字化技术的发展和 3D 打印技术的成熟而演变的。

随着社会的进步和生活水平的提高，社会对产品多样化、低制造成本及短制造周期等需求日趋迫切，传统的制造技术已不能满足市场多品种、个性化的产品需求。因此，过去是企业越大越有竞争力，今后是越小越有竞争力。小到什么程度呢？小到个人！每个人都能在网络社区里提供一个设计方案，从首饰、服装到手表、汽车，设计已不再神秘。虽然初始的设计很原始、很粗糙，但在互联网上有众人的参与改进，并各自加入自己独特的新思路，以此形成各式各样的产品变种，可分别满足某部分特定偏好的人群。比如，有的设计者把产品的客户定位为 21 岁的年轻女性群体，另外一些设计者主要面向 65 岁的老年男性群体，因此各自都需要对产品设计做个性化的创新。

这些高度个性化的产品从设计变成现实，需要技术的支撑，智能数字化和 3D 打印则是重要的技术手段。因为这种追求高附加值的**个性化定制**，之前都是以较大的手工工作量为代价的，尤其是当需要大规模“**批量定制**”（Mass Customization，大规模定制）时。比如，需要为一万名用户定制个性化的眼镜、服装、帽子、鞋子，如果使用人工逐一为每位用户进行手工测量和手工设计，工作量和成本都将变得不可接受。因此，为提高定制效率，智能数字化技术将发挥关键的作用。比如，可利用摄像头自动采集、分析提取每位用户的体貌个性特征，并自动根据视觉美

感进行形状设计、颜色肤色搭配等，可极大地缩减定制周期。换言之，3D 智能数字化技术是 3D 打印实现“低成本大规模定制”、以区别于传统“昂贵费时的手工化定制”的基础和关键所在。

在工业制造领域，这种灵活的生产方式被称为“柔性制造”（Flexible Manufacturing），其价值在于高效率地应对市场需求的多样性和变化。“柔性”是相对于“刚性”而言的。以福特汽车公司的大规模流水线生产为代表，传统的“刚性”自动化生产线主要实现单一品种的大批量生产。其优点是生产效率很高，由于设备是固定的，所以设备利用率也很高，单件产品的成本低。但只能加工一种或几种相类似的零件，难以应付多品种、个性化定制的生产。

之后，为了提高生产过程的柔性，逐渐转变为以丰田汽车公司为代表的以模块化为核心的精益生产（Lean Production）。精益生产将整个生产环节拆分，模块之间技术关系被相对固定下来，但模块内部可以存在多种变化，以变化组合的方式来满足需求的多样化。由于模块之间的技术关系相对固定，模块组件之间的组装可以实现较高程度的标准化和规模经济。同时，由于单一模块内部的创新，不会影响模块间的技术关系，又可以给最终商品带来性能上的改变。因此，相对于福特标准化的生产方式，局部创新在精益生产条件下更容易实现。

虽然以丰田汽车为代表的精益生产相对于福特标准化生产的柔性程度有明显的提升，但这种模块化的生产方式还不够充分柔性，尤其当人们越来越多地通过追求定制化的商品来彰显自己的个性时。3D 打印设备的出现在很大程度上消除了这些局限性，这归功于 3D 打印技术的特点：可以在成本几乎不变的条件下，实现形状的任意变化。最具革命的意义在于，在一些制造领域，3D 打印技术无须使用模块组件，通过一次成型即可直接实现最终产品的多样化，从而颠覆精益生产，走向自由制造（Freedom Fabrication）的生产阶段。3D 打印技术的智能数字化设计、快速成型制造的工艺特点，可以大量节省研发者制造的时间，加速设计创新实现的速度。

柔性制造实现的关键在于，智能数字化技术在控制、检测、监控和 3D 仿真等方面得到广泛应用，涉及的学科领域包括模式识别与智能系统、计算机视觉、3D 计算机图形学、自动化控制等。届时，智能化机械与人之间将相互融合，柔性地全面协调从接受订单至生产、销售等企业生产经营的全部活动。智能制造技术（IMT, Intelligent Manufacturing Technology）应运而生，其旨在将智能数字化融入制造过程的各个环节，借助模拟专家的智能活动，取代或延伸制造环境中人的部分脑力劳动。在制造过程中，系统能自动监测其运行状态，在受到外界或内部激励时能自动调节其参数，以达到最佳工作状态，具备自组织能力。因此 IMT 被称为 21 世纪的制造技术。此外，对未来智能化柔性制造技术具有重要意义且正在急速发展的一个领域是智能传感器技术。该项技术是伴随计算机应用技术和人工智能而产生的，它使传感器具有内在的“决策”功能。同时，随着摄像机硬件技术的不断完善，视觉计算方法的不断发展，使用视觉系统提高柔性和精度将成为今后制造业的研究热点。

小结一下，如何开发出能够激发用户个性化创造热情的“杀手级软件应用”，让广大用户能够“随时随地”创建出各种各样的 3D 数字化形状，将是引爆 3D 打印“无所不在地”（ubiquitously）走进千万家庭的关键之一。因此在本书后续章节中（如第 6 章），将对构造个性化形状的各种 3D 智能数字化方法进行详细阐述，包括个性化特征检测与匹配、个性化形状建模与扫描、个性化形状编辑与合成、个性化形状分析与处理、个性化形状检索等。

4.5 智能云网：云端智能服务和云制造

智能数字化技术涉及视觉计算、模式识别与智能系统、复杂系统与自动控制、数据挖掘与机器学习等众多“高科技”学科，普通技术人员掌握的门槛很高。因此，这些技术将来会以云端智能化服务的形式提供给普通用户和开发者。以定制一双鞋子为例，普通用户只需在手机上下载一个 App 应用，给自己的双脚拍几张照片，并指定喜欢的款式和颜色，之后位于云端的智能计算服务将根据用户上传的照片重建出 3D 脚型，然后把鞋子设计出来。所涉及的复杂智能技术全都在云端完成，App 的开发者根本无须了解。同样，用户也无须了解专业化的制造技术，只需轻轻点击提交订单后，系统就会自动在云制造集群中搜索到邻近的打印节点，快速打印后马上就可以送货上门。在这一商业模式中，与传统制造企业不同的是：App 开发者和用户都无须厂房建设、人员雇佣等大量的前期投入，却可实现之前不敢想象的个人智造。

我们把 3D 打印产业模式所依赖的云端智能化服务和云制造统称为“**智能云网**”（ICN, Intelligent Cloud Network），其具体的运作流程请参见第 1 章第 1.4.3 节的图 1-28。智能化、云端化、网络化、数字化将是 3D 打印未来的重要特点。智能云网在虚拟化的数字设计和现实的物质世界中建立了一个通用的智能转化平台，改变了之前受制于专业设计技术和专业加工技术的高壁垒而无法在物质世界中实现个人智造的窘境。在这一过程中：想法和创新真正做到了与专业理论知识、高级加工技能、制造设备的分离，使得普通个人即便在缺乏深厚专业知识和熟练制作加工技能的情况下，仍然可以创新性地将自己的想法变成实物。



提示：苹果公司的前 CEO 史蒂夫·乔布斯（Steve Jobs）是个典型的想法和创意天才。然而，可能他自己也不会承认自己是个技术天才。但这并不妨碍他设计出像 iPod、iPhone 这样的划时代产品，因为他并不需要关心像多进程操作系统运行原理、Siri 语音识别原理、CPU 芯片加工制造等技术细节。同样地，3D 打印和智能云网的出现，让每一个有想法和创意的个人，都有可能成为某个细分领域的史蒂夫·乔布斯。

以上涉及云制造的概念，其对 3D 打印这种“规模定制”的运营模式尤其关键。维基百科对云制造（Cloud Manufacture）的定义是：“具有各种制造资源和能力，可以智能检测并连接更广泛的互联网，具备自动管理和控制能力”。每个单独的制造节点都是自主的、通过网络互联的。云制造的优点是资源可以扩展，还可自动平衡负载。制造商可以根据项目的特别需求，如本批次是定制一千件还是一万件，来构建一个临时的集群。每个云制造商的产能可能很小，但集群后的整体产能完全可以满足项目需求，且非常经济、灵活。

3D 打印的革命意义不在于替代规模制造，而在于成就个性化的生产模式。和标准化的流水线制造相比，3D 打印在一段时间内还不具备规模生产的经济性。但当这一技术将专业技术封闭的制造大门向普通个体打开时，你无法想象，它能激发出多少个性化的创意设计。而在网络交易平台和云制造的辅助下，这些创意转化为可以赢利的产品，并将以几何级数的速度增长。单个个性化设计产品，或许只能满足小众市场，但却是标准化生产无法胜任的。未来制造业的创新主体将从专业化的企业转变为小微组织，或自然个体。

“个人智造”的商业模式已经拉开序幕。2011 年，以 3D 打印服务和产品交易为主要功能的网络平台 Shapeways，产品销售量实现约 1000% 的增长！一个“个人智造”领域的 C2C Consumer

to Consumer，个人到个人）模式似乎已跃跃欲试。由于不受制于专业制造标准的限制，这种个人生产模式的创新速度可能远高于专业制造企业。同期，全球个人 3D 打印设备销售量同比增长接近 300%，大幅超越工业级 3D 打印机的增速。这些数据似乎在佐证个体创意通过 3D 打印追求自我实现的欲望正在急速膨胀。而个人 3D 打印作为一种可在产品设计阶段初步验证想法可行性的工具，有望再现之前个人电脑走过的辉煌征程，成为广泛普及的大众消费品。

4.6 大数据和深度学习：3D打印内容的挖掘与推荐

有了 3D 打印机，那么我们日常究竟要打印些什么东西呢？这就像我们有了 MP3 播放器 iPod，却在为每天要听什么歌而犯愁。因为，现在网上可下载的 MP3 实在太多了，数不胜数。同样，目前网上的 3D 模型也数不胜数，今天，你到底应该打印哪些模型呢？

造成你如此困惑的根源就在于“大数据”。

4.6.1 什么是大数据

“大数据”（Big Data）是“数据化”趋势下的必然产物。数据化带来了两个重大的变化。一是数据量的爆炸性剧增，最近几年所产生的数据量等同于 2010 年以前整个人类文明产生的数据量总和。以前网上的 3D 模型非常少，而目前仅 Shapeways 这一个网站上的 3D 模型，就已突破了 100 万个。二是数据来源的多样化以及**异构性**，比如介绍某款手机产品的网页，既有文本、语音，还有视频、图像、3D 模型等，从各个方面展示了该产品的特征，这种多源性也有助于滤除数据噪声、交叉验证。数据间是否具有结构性和关联性，是“**大数据**”与“**大规模数据**”的重要差别；“大数据”这一概念中包含着对数据对象的处理行为，即快速挖掘和展现其中蕴含着的有价值信息。

大数据的特点可总结为 4 个“V”——Volume（**体量巨大**）、Variety（**类型多样**）、Value（**价值密度低，商业价值高**）、Velocity（**处理速度快**）。牛津大学互联网研究所维克托·迈尔·舍恩伯格教授指出，“大数据”所代表的是当今网络社会所独有的一种新型能力——通过对海量数据进行分析，来获得有巨大价值的产品和服务或深刻的洞见。例如，你在网上买书时，网站根据你之前的购买记录快速推测你的阅读类型（比如你喜欢魔幻武侠小说），然后把当前最热门的 3 部魔幻武侠小说显示在网页最醒目的位置，以便激发你的购买欲。因此，可利用大数据对客户群进行细分，通过分析其既往行为，推测他们潜在的意图、习惯和计划，以实现**精准营销**。

大数据时代会颠覆许多传统思维，在哲学层面体现为“**经验主义**”比“**理性主义**”更多地被人们所采用。以前人们总在探寻问题的因果：事物为什么会这样？但现在，人们更关心结论。比如，从大量数据分析得出冬天第一场雪过后大白菜价格会涨大概两倍，那么商家会更乐意利用这个结论来关注天气预报并伺机囤积大白菜，而不会像专家那样坐在一起讨论为什么第一场冬雪后大白菜会涨价、为什么是涨两倍而不是涨 3.2 倍。大数据也意味着对效率的追求，而不是去过分追求数值上的精确。

专家的价值在于因果分析，而大数据却放弃对因果关系（Causality）的追求，仅关注相关关系（Correlation）。也就是说，只需要知道“**是什么**”，而不需要知道“**为什么**”。这种变化已经远远突破了技术层面，将对人类认识世界的哲学观产生重大影响。因果关系只是相关关系中特

殊的一种，大数据告诉我们很多情况下只要关注相关关系以做出预测就够了。另一种可能的解释是，数据是不会骗人的，而人（即使是专家）的见解往往是主观和偏见的。当然，我们并不是说逻辑性的因果关系不重要，而是我们一开始往往会迷失在纷乱繁杂的数据海洋中、毫无头绪，所以这时就可首先想办法获得**统计意义上的相关关系**，然后再考虑从中提取出**逻辑性的因果关系**。这其实很好理解：当我们对数据无法直接获得可解释性时，那就试着先观察出这些数据的统计规律性（“是什么”），然后再针对这些规律进行解释（“为什么”）。

大数据还有一个巨大的优势是，可利用**通用的统计学模型**代替各种各样的**专家系统**，“以不变应万变”。例如，基于大数据（包罗万象的语料数据），Google 的翻译算法可统一实现几十种语言（英语、汉语、法语、韩语、拉丁语等）的互译，而无须针对每种语言定制专门的语法专家系统。IBM 公司的 Fred Jelinek 院士是利用大数据进行统计语音识别与合成的著名学者，他曾说过一句著名的论点：“每当我解雇一个语言学家，语音识别系统的性能就会改善一些”。

大数据是网络社会在掌握海量数据收集、存储和处理技术基础上所产生的一种进行判断和预测的能力。专家往往希望归纳出一个**模型**，而在大数据时代，**数据**直接自己“说话”，变得比模型更重要，因为再复杂的模型也无法包罗万象。而当数据“大”（多）到能对几乎整个样本空间进行充分覆盖时，就可以减弱对理论和模型的依赖，不再需要通过模型去经历“从**特殊**归纳（**Induce**）到**一般**，再从**一般**演绎（**Deduce**）到**特殊**”的传统流程，而是利用大数据去直接实现“从**特殊到特殊**”的判断和预测（这种直接的方式也被称为**转导，Transduce**），因为大数据中已经包含了足够多的“特殊”样本以供参考。换言之，此时数据本身便是模型，也即大数据可实现**全样**而非**抽样**（现实中要获得代表真实情况的抽样非常难，比如可能会因为抽样不够全面而遭遇“黑天鹅事件”）。

大数据将给整个社会带来从生活到思维上革命性的变化：人们所接受的服务，将以数字化和个性化的方式呈现，借助 3D 打印技术和智能数字化，零售业和医疗业也将实现数字化和个性化的服务。



扩展：除了大数据，还有所谓的小数据（iData）。小数据跟大数据的根本区别在于：小数据以单个人（**个体**）为唯一对象，重点在于**深度**，即像一位忠诚细致的“个人管家”那样对个人数据进行全方位、全天候地深入精确分析，同时还可主动灵活地设置各种外界访问权限以保护个人隐私；而大数据则侧重于在某个领域（**群体**），大范围、大规模地进行数据的全面收集处理分析，侧重点在于**广度**。

目前，**Hadoop** 是最为流行的大数据处理平台，是一个开源的、可运行于大规模集群上的分布式并行编程框架，由分布式文件系统（如 HDFS）、数据库（如 HBase，属于 NoSQL 类型的数据库）、数据处理模块（如分布式编程模型 **MapReduce**）等组成。借助于 Hadoop，程序员可以轻松地编写分布式并行程序，将其运行于大规模集群上，从而完成大数据的计算。除了 Hadoop，此外还有另一个高效的分布式并行计算系统 **Spark**，通用性更好、迭代运算效率更高、容错能力更强，目前其发展势头正逐渐盖过 Hadoop。



扩展：数据挖掘不仅与统计学习有关，而且与信息论紧密相关。所谓信息，根据信息论创始人**香农**（Claude Elwood **Shannon**）的说法：“凡是在一种情况下能**减少不确定性**的任何事物都叫作**信息**”。在信息论中，使用**信息熵**（**Entropy**、**Shannon Entropy**，

简称：熵）来评估信息量的大小，即**不确定性的度量**：

$$H(\mathbf{X}) = -\sum_{x \in \mathbf{X}} p(x) \log p(x)$$

通过上式可以看出信息熵被定义为信息（ $-\log p(x)$ ）的期望值，**单位为比特（bit）**。事件的**不确定性越大**，则**信息熵就越大**（也即**把它搞清楚所需的信息量就越大**）。比如，“人咬狗”相比于“狗咬人”是小概率事件，可能性小，不确定性大，因此熵更大。

条件熵（Conditional Entropy）的定义：

$$H(\mathbf{X}|\mathbf{Y}) = -\sum_{x \in \mathbf{X}, y \in \mathbf{Y}} p(x, y) \log p(x|y)$$

可证明 $H(\mathbf{X}) \geq H(\mathbf{X}|\mathbf{Y})$ ，也即如果增加了（与 \mathbf{X} 相关的） \mathbf{Y} 的信息， \mathbf{X} 的不确定性下降了。类似地，还有 $H(\mathbf{X}|\mathbf{Y}) \geq H(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$ 。

那么， \mathbf{X} 与 \mathbf{Y} 到底有多相关呢？我们可通过**互信息（Mutual Information）**来量化地度量“相关性”：

$$I(\mathbf{X}; \mathbf{Y}) = \sum_{x \in \mathbf{X}, y \in \mathbf{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y})$$

比如，“计算机”和“鼠标”这两个词的互信息就比“计算机”和“牙刷”的互信息更大，因为前者更相关。

相对熵（Relative Entropy），又叫 **KL 距离（Kullback-Leibler Divergence，KL 散度）**、**信息增益（Information Gain）**、**信息散度（Information Divergence）**：

$$KL(f(x) \| g(x)) = \sum_{x \in \mathbf{X}} f(x) \cdot \log \frac{f(x)}{g(x)}$$

不同于前面的熵和互信息（它们衡量的是随机变量的关系），相对熵衡量的是两个概率分布函数的差异程度。

4.6.2 大数据背景下的个性化推荐系统

我们回到一开始的话题，总有一天你肯定会为每天要打印哪些 3D 模型而犯愁，因为网上的模型实在是太多了。答案是：**个性化推荐系统**！基于大数据（比如你以往所有的打印记录、你好友圈的所有打印记录，还有网上所有的 3D 模型），推荐系统对你的个性偏好进行深度分析（结合 4.6.3 节将要介绍的深度学习算法）确认，发现你偏好于打印可爱类型的美少女模型。而且通过对模型形状特点（参见第 6 章的 6.2 节）的无监督特征学习（参见 4.6.3 节）发现：你只喜欢拥有大酒糟鼻子的脸型（您的一个特殊癖好）。于是，推荐系统就把那些近期刚刚上架不久的热门美少女模型推荐给你，甚至还帮你自动分类（参见第 6 章的 6.4.1 节），设计成多个主题系列，让你摇身变成了一位有高端品位的“卡哇伊大酒糟鼻子美少女 3D 模型”收藏专家。

言归正传，在大数据的社会背景下，所谓个性化推荐就是根据用户的兴趣特点和购买行为，

向用户推荐感兴趣的信息和商品。在当前 Web 2.0 时代，随着电子商务规模的不断扩大，商品数量和种类快速增长，顾客需要花费大量的时间才能找到自己想买的商品，出现了所谓的**信息超载**（Information Overload）问题。为了解决这个难题，**个性化**（Personalized）推荐系统（Recommender System）应运而生。个性化推荐系统通过分析用户的行为，发现用户的个性化需求、兴趣等，然后将用户感兴趣的信息、产品推荐给用户。推荐系统主要依赖于数据挖掘（Data Mining）和机器学习（Machine Learning）的算法。

通常，以 Google、百度为代表的搜索引擎可以让用户通过输入关键词精确找到自己需要的相关信息。但是，如果用户无法想到准确描述自己需求的关键词，此时搜索引擎就无能为力了。和搜索引擎不同，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为来对用户的兴趣进行建模，从而主动给用户推荐可满足他们兴趣和需求的信息。因此，搜索引擎和推荐系统对用户来说是两个互补的工具，前者需要用户“主动出击”，后者则让用户“被动笑纳”。

推荐系统现已广泛应用于很多领域，其中最典型并具有良好的发展应用前景的领域就是电子商务领域，如图 4-41 所示的亚马逊购物网站。



图 4-41 亚马逊推荐系统的用户界面

推荐系统可认为是一种特殊形式的**信息过滤**（Information Filtering）系统，主要有“协同过滤推荐”、“基于内容的推荐”、“基于关联规则的推荐”、“基于知识推理的推荐”、“组合推荐”这几种智能算法。

如果推荐系统根据用户的历史兴趣来给用户做推荐，那么这种方法被称为“**协同过滤推荐**”（Collaborative Filtering Recommendation）算法。协同过滤是基于这样的原理：首先找到与此用户有相似兴趣的其他用户，然后将他们感兴趣的内容推荐给此用户。其基本思想非常易于理解，在日常生活中，我们往往会通过好朋友的推荐来进行一些选择，如音乐、电影等。协同过滤实际上是通过人与人之间的合作来过滤掉不良信息，因此协同过滤也叫**社会过滤**（Social Collaborative Filtering）。



参考：具体地，给定用户 A 和 B ，令 $S(A)$ 是 A 喜欢的物品集合， $S(B)$ 是 B 喜欢的物品集合，则用户 A 和 B 的兴趣相似度可用如下公式来表示：

$$sim_{AB} = \frac{|S(A) \cap S(B)|}{|S(A) \cup S(B)|} \text{ (Jaccard 公式) 或 } sim_{AB} = \frac{|S(A) \cap S(B)|}{\sqrt{|S(A)| |S(B)|}} \text{ (余弦相似度)}$$

然后，用户 A 对物品 i 的喜欢程度就可用如下公式来计算：

$$like_{Ai} = \sum_{B \in S(A, k) \cap S(i)} sim_{AB} like_{Bi}$$

其中， $S(A, k)$ 包含与用户 A 兴趣最接近的 k 个用户， $S(i)$ 为喜欢物品 i 的用户集合。

如果推荐系统利用了商品的内容描述，计算用户的兴趣和商品描述之间的相似度，来给用户做推荐，则称为“**基于内容的推荐**”（Content-based Recommendation）算法，如 4.6.1 节提到的魔幻武侠小说的推荐例子。基于内容的推荐不需要其他用户的帮助，不需要依据其他用户对商品的评价意见，因此没有冷启动问题、新上架商品问题、冷门商品问题和**稀疏性问题**（相比于海量的商品数目，商品的 用户评价数目往往非常少，其本质上可归结为数据的高维）。说直白点，就是商品刚上架时，此时还没有一个用户对该商品做出过评价，这时“协同过滤推荐”算法就无法工作，而“基于内容的推荐”算法则没有问题。

基于关联规则的推荐以关联规则为基础，分析用户已经选择的项目与未选择项目之间的关联性得出最后的推荐结果。关联规则推荐的典型例子是购物车分析，该推荐方法通过发现顾客放入其购物车中不同商品之间的联系，分析顾客的购买习惯。例如买面包的顾客，还会购买牛奶。通过了解哪些商品频繁地被顾客同时购买，可以把相关的产品摆在一起，达到促销的目的。

基于知识推理的推荐是数据挖掘技术在个性化推荐系统中的应用，它不参考用户对于项目的偏好，而是依据某种知识或者推理来进行推荐。例如，如果用户喜欢冲洗大照片，那么高分辨率相机会对其更有吸引力。具体地，通过对数据库中的数据分析，发现信息中隐含的有价值的知识，或者在用户和待推荐对象之间构建推理，来进行推荐。此外，基于知识推理的推荐系统还可不断学习用户对推荐的反馈情况，从而达到更高的推荐质量。

除以上这些，还有基于信任网络的推荐系统、上下文感知推荐系统、基于网络结构的推荐等。由于各种推荐方法都有优缺点，所以在实际中，**组合推荐**（Hybrid Recommendation）经常被采用。比如分别用基于内容推荐方法和协同过滤推荐方法去产生一个推荐预测结果，然后用某方法组合其结果，以弥补各推荐技术的弱点。判断一个推荐系统的优劣主要有以下**评价指标**：准确度（Accuracy，包括预测准确度、分类准确度、排序准确度）、覆盖率（Coverage）、惊喜性（Serendipity）、新颖性（Novelty）、**多样性、用户满意度**等。

推荐系统的早期研究主要集中在静态用户行为分析领域，即不考虑用户行为发生的时间，而仅仅研究用户行为中与时间无关的静态模式。近年来，很多研究人员转向研究推荐系统的**动态特性**，主要包括用户兴趣变化的动态模型，基于时间上下文的推荐等问题。例如一部电影刚上映的时候可能会被很多人关注，但过了几个月后人们逐渐不再感兴趣，所以这时就不能再把它放在最醒目的位置，即使这部电影的评价很高、用户的兴趣跟它很相关。

推荐系统可以更好地发掘信息的**长尾 (Long Tail)**。在传统零售超市里，最热门的少数商品往往摆在最醒目的位置，而大量的冷门商品则放在货架的某个角落，很难让人注意到。但在电子商务时代，借助于个性化推荐系统，这些冷门商品也终于可以扬眉吐气、主动被推送到感兴趣用户网页的最醒目位置。

4.6.3 深度学习：像人脑一样深层次地思考

从 4.6.2 节我们可以看出，个性化推荐系统确实很会“察言观色”，针对不同的用户，主动推送不同的 3D 打印内容。但如果你认为它真正有了“人工智能”，那你就错了。其实，这些推荐系统背后的运行原理主要基于概率统计、矩阵或图模型，计算机对这些数值运算确实很擅长，但由于采用的只是“**经验主义**”的实用方法（也即管用就行），而非以“**理性主义**”的原则真正探求智能产生的原理，所以距离真正的人工智能还很远。**AI (Artificial Intelligence)**，也就是**人工智能**，就像长生不老和星际漫游一样，是人类最美好的梦想之一。虽然计算机技术已经取得了长足的进步，但是到目前为止，还没有一台计算机能产生“自我”的意识。



提示：图灵测试 (Turing Testing)，是计算机是否真正具有人工智能的试金石。“计算机科学之父”及“人工智能之父”英国数学家阿兰·图灵（1912—1954）在 1950 年的一篇著名论文《机器会思考吗？》里，提出图灵测试的设想。即把一个人和一台计算机分别隔离在两间屋子，然后让屋外的一个提问者对两者进行问答测试。如果提问者无法判断哪边是人，哪边是机器，那就证明计算机已具备人的智能。

直到**深度学习 (Deep Learning)** 的出现，让人们看到了一丝曙光，至少，（表象意义下的）图灵测试已不再是那么遥不可及了。2013 年 4 月，《麻省理工学院技术评论》杂志将深度学习列为 2013 年十大突破性技术（Breakthrough Technology）之首。有了深度学习，推荐系统可以更加深度地挖掘你内心的需求，并从海量的 3D 模型库中挑选出最合适的供你打印。

让我们先来看看人类的大脑是如何工作的。1981 年的诺贝尔医学奖，颁发给了 David Hubel 和 Torsten Wiesel，以及 Roger Sperry。前两位的主要贡献是，发现了人的视觉系统的信息处理是分级的。如图 4-42 所示，从视网膜（Retina）出发，经过低级的 V1 区提取边缘特征，到 V2 区的基本形状或目标的局部，再到高层的整个目标（如判定为一张人脸），以及到更高层的 PFC（前额叶皮层）进行分类判断等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表达越来越抽象和概念化，也即越来越能表现语义或者意图。

这个发现激发了人们对于神经系统的进一步思考。大脑的工作过程，或许是一个不断迭代、不断抽象概念化的过程，如图 4-43 所示。例如，从原始信号摄入开始（瞳孔摄入像素），接着做初步处理（大脑皮层某些细胞发现边缘和方向），然后抽象（大脑判定眼前物体的形状，比如是椭圆形的），然后进一步抽象（大脑进一步判定该物体是张人脸），最后识别眼前的这个人——正是大明星刘德华。这个过程其实和我们的常识是相吻合的，因为复杂的图形，往往就是由一些基本结构组合而成的。同时我们还可以看出：**大脑是一个深度架构**，认知过程也是深度的。

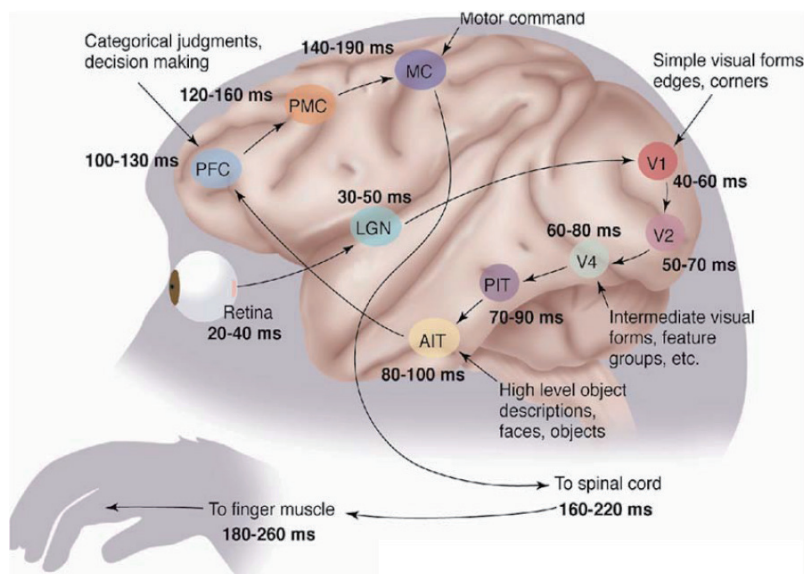


图 4-42 人脑的视觉处理系统(图片来源 : Simon Thorpe)

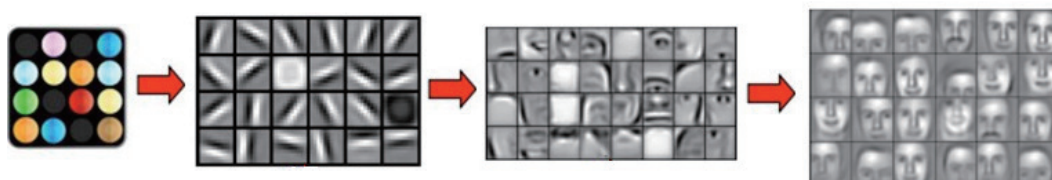


图 4-43 视觉的分层处理结构(图片来源 : Stanford)

而深度学习 (Deep Learning), 恰恰就是通过组合低层特征形成更加抽象的高层特征 (或属性类别)。例如, 在计算机视觉领域, 深度学习算法从原始图像去学习得到一个低层次表达, 例如边缘检测器、小波滤波器等, 然后在这些低层次表达的基础上, 通过线性或者非线性组合, 来获得一个高层次的表达。此外, 不仅图像存在这个规律, 声音也是类似的。比如, 研究人员从某个声音库中通过算法自动发现了 20 种基本的声音结构, 其余的声音都可以由这 20 种基本结构来合成!

在进一步阐述深度学习之前, 我们需要了解什么是**机器学习 (Machine Learning)**。机器学习是人工智能的一个分支, 而在很多时候, 几乎成为人工智能的代名词。简单来说, 机器学习就是通过算法, 使得机器能从大量历史数据中学习规律, 从而对新的样本做智能识别或对未来做预测。

而深度学习又是机器学习研究中的一个新的领域, 其动机在于建立可以模拟人脑进行分析学习的神经网络, 它模仿人脑的机制来解释数据, 例如, 图像、声音和文本。深度学习之所以被称为“深度”, 是因为之前的机器学习方法都是浅层学习。深度学习可以简单理解为**传统神经网络 (Neural Network)**的发展。大约二三十年前, 神经网络曾经是机器学习领域特别热门的一个方向, 这种基于统计的机器学习方法比起过去基于**人工规则**的**专家系统**, 在很多方面显示出

优越性。如图 4-44 所示，深度学习与传统的神经网络之间有相同的地方，采用了与神经网络相似的分层结构：系统是一个包括输入层、隐层（可单层、可多层）、输出层的多层网络，只有相邻层节点（单元）之间有连接，而同一层以及跨层节点之间相互无连接。这种分层结构，比较接近人类大脑的结构（但不得不说，实际上相差还是很远的，考虑到人脑是个异常复杂的结构，很多机理我们目前都是未知的）。

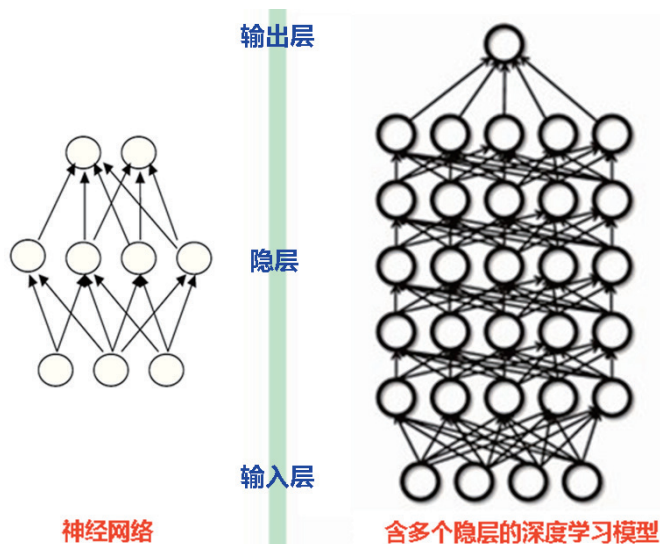


图 4-44 传统的神经网络与深度神经网络



提示：人类大脑由千亿个神经元组成，同时每个神经元平均连接到其他几千个神经元，这样形成一个庞大的神经元网络。通过这种连接方式，神经元可以收发不同数量的能量，但它们对能量的接受并不是立即做出响应，而是先累加起来，只有当累加的总和达到某个临界阈值时才把能量发送给其他的神经元。而**人工神经网络（Artificial Neural Networks, ANN）**将**人类神经网络**做了数学上的抽象，如图 4-47 所示，将其抽象为输入层、输出层以及中间的若干**隐层（Hidden Layer）**，用于层次化地对内在特征进行降维和抽象表达，相当于特征检测器），其中每层都有若干节点及连接这些点的边，通过在训练数据集上学习算出边的**权重（Weight）**来建立模型。边所表征的函数（通常为非线性函数）的不同，对应于不同的神经网络。例如，第 6 章 6.4.1 节所介绍的感知机就是一种最简单的、不含任何隐层的**前向（Feedforward）人工神经网络**，其中的函数 $w \cdot x + b$ 被称为**传递函数（Transfer Function）**，而门限截止函数 $\text{sign}()$ 则被用作**激活函数（Activation Function）**。在 20 世纪七八十年代，这种在人工智能领域被称为联结主义学派（Connectionism）的方法曾盛极一时。

但是后来，因为理论分析的难度，加上训练方法需要很多经验和技巧，以及巨大的计算量和优化求解难度，神经网络慢慢淡出了科研领域的主流方向。值得指出的是，神经网络（如采用误差反向传播算法：Back Propagation，简称 BP 算法，通过**梯度下降**方法在训练过程中修正权重使得网络误差最小）在层次深的情況下性能变得很不理想（传播时容易出现所谓的梯度弥散 Gradient Diffusion 或称之为梯度消失，根源在于非凸目标代价函数导致求解陷入**局部最优**，且这

种情况随着网络层数的增加而更加严重，即随着梯度的逐层不断消散导致其对网络权重调整的作用越来越小)，所以只能转而处理浅层结构（小于等于3），从而限制了性能。于是，20世纪90年代，有更多各式各样的浅层模型相继被提出，比如只有一层隐层节点的支撑向量机（SVM，Support Vector Machine）和 Boosting，以及没有隐层节点的最大熵方法（例如 LR，Logistic Regression）等，在很多应用领域取代了传统的神经网络。

显然，这些浅层结构算法有很多局限性：在有限样本和计算单元情况下对复杂函数的表示能力有限，针对复杂分类问题其泛化能力受到一定的制约。更重要的是，浅层模型有一个特点，就是需要依靠人工来抽取样本的特征。然而，手工地选取特征是一件非常费力的事情，能不能选取好很大程度上靠经验和运气。既然手工选取特征不太好，那么能不能自动地学习一些特征呢？



提示：实际生活中，人们为了实现对象的分类，首先必须做的事情是如何来表达一个对象，即必须抽取一些**特征**来表示一个对象。例如，区分人和猴子的一个重要特征是有无尾巴。特征选取的好坏对最终结果的影响非常大。

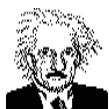
此外，我们希望提取到的特征能代表输入数据的最重要部分，就像 PCA（Principal Component Analysis，主成分分析，请参见第6章的6.2.2节）那样，找到可以代表原信息的主要成分。以**自动编码器（AutoEncoder）**为例，这是一种尽可能复现输入信号的神经网络：即输出 y 要尽可能与输入 x 相同，表示为 $\min \|x - y\|$ 。我们可通过训练调整这个神经网络的参数，来得到每一层中的权值系数，这样就可得到输入 x 的一个层次化的表示。这个可代表原信息主要成分的表示就是所谓的特征。

进一步地，我们还可用 $y = Wh$ 来表示输出 y ，其中 W 称为**字典**。类似于 PCA， W 可理解为基， h 可理解为系数。同时，我们不仅希望将信号表示为一组层次化基的线性组合，而且要求**只需较少**的几个基就可以将信号表示出来，这就是所谓的**稀疏编码（Sparse Coding）**。“稀疏性”定义为：只有很少的几个非零元素或只有很少的几个远大于零的元素。也即，我们希望求得一组最佳的系数 h^* ，满足：

$$h^* = f(x) = \arg \min_h \|x - Wh\|_2^2 + \lambda \|h\|$$

注意上式右边对系数采用了**L1 范式 / 正则化 / 约束**以满足稀疏性，上式实际上是对**Lasso（The Least Absolute Shrinkage and Selectionator operator）**估计的求解。之所以希望“**稀疏性**”是科学依据的，因为绝大多数的感官数据，比如自然图像，都可以被表示成“**少量**”基本元素的叠加，比如基本线 / 面的叠加。稀疏编码算法是一种无监督学习方法，它用来寻找一组“**超完备**”基向量（基向量的个数比输入向量的维数要大）以更高效地表示样本数据，以找出隐含在输入数据内部的结构与模式。

答案是能！深度学习框架将特征和分类器结合到一个框架中，**自动地从海量大数据中去学习特征，在使用中减少了手工设计特征的巨大工作量**。看它的一个别名：**无监督特征学习（Unsupervised Feature Learning）**，就可以顾名思义了。**无监督（Unsupervised）学习**的意思就是不需要通过人工方式进行样本类别的标注来完成学习。因此，深度学习是一种可以自动地学习特征的方法。



提示：准确地说，深度学习首先利用**无监督学习**对每一层进行逐层预训练（Layerwise Pre-Training）**去学习特征**；每次单独训练一层，并将训练结果作为更高一层的输入；然后到最上层改用**监督学习**从上到下进行微调（Fine-Tune）**去学习模型**。

深度学习通过学习一种深层非线性网络结构，只需简单的网络结构即可实现复杂函数的逼近，并展现了强大的从大量无标注样本集中学习数据集本质特征的能力。深度学习能够获得可更好地表示数据的特征，同时由于模型的层次深（通常有 5 层、6 层，甚至 10 多层的隐层节点，“深”的好处是可以控制隐层节点的数目为输入节点数目的多项式倍而非多达指数倍）、表达能力强，因此有能力表示大规模数据。对于图像、语音这种特征不明显（需要手工设计且很多没有直观的物理含义）的问题，深度模型能够在大规模训练数据上取得更好的效果。尤其是在语音识别方面，深度学习使得错误率下降了大约 30%，取得了显著的进步。相比于传统的神经网络，深度神经网络做出了重大的改进，在训练上的难度（如梯度弥散问题）可以通过“逐层预训练”来有效降低。注意，深度学习不是万金油，像很多其他方法一样，它需要结合特定领域的先验知识，需要和其他模型结合才能得到最好的结果。当然，还少不了需要针对自己的项目去仔细地调参数，这也往往令人诟病。此外，类似于神经网络，深度学习的另一局限性是可解释性不强，像个“黑箱子”一样不知为什么能取得好的效果，以及不知如何有针对性地去具体改进，而这有可能成为产品升级过程中的阻碍。

深度学习通过很多数学和工程技巧增加（堆栈叠加：Stack）隐层的层数，如果隐层足够多（也就是深），选择适当的连接函数和架构，就能获得很强的表达能力。**深度学习的一个主要优势在于可以利用海量训练数据（即大数据）**，但是常用的模型训练算法反向传播（Back Propagation）仍然对计算量有很高的要求。而近年来，得益于计算机速度的提升、基于 MapReduce 的大规模集群技术的兴起、GPU 的应用以及众多优化算法的出现，耗时数月的训练过程可缩短为数天甚至数小时，深度学习才在实践中有了用武之地。

值得一提的是，深度学习的诞生并非一帆风顺。虽然 Yahn Lecun 在 1993 年提出的**卷积神经网络（CNN，Convolutional Neural Network）**是第一个真正成功训练多层网络结构的学习算法，但应用效果一直欠佳。直到 2006 年，Geoffrey Hinton 基于**深度置信网（DBN，Deep Belief Net）**——其由一系列**受限波尔兹曼机（RBM，Restricted Boltzmann Machine）**组成，提出非监督贪心逐层训练（Layerwise Pre-Training）算法，应用效果才取得突破性进展，其与之后 Ruslan Salakhutdinov 提出的**深度波尔兹曼机（DBM，Deep Boltzmann Machine）**重新点燃了人工智能领域对于**神经网络（Neural Network）**和**波尔兹曼机（Boltzmann Machine）**的热情，才由此掀起了深度学习的浪潮。从目前的最新研究进展来看，只要数据足够大、隐层足够深，即便不加“Pre-Training”预处理，深度学习也可以取得很好的结果，反映了大数据和深度学习相辅相成的内在联系。此外，虽说非监督（如 DBM 方法）是深度学习的一个优势，深度学习当然也可用于带监督的情况（也即给予了用户手动标注的机会），实际上带监督的 CNN 方法目前应用得越来越多，乃至正在超越 DBM。



提示：与前向神经网络不同，RBM（**受限波尔兹曼机**）中的可见层和隐含层之间的连接是无方向性且全连接的。**对比差异无监督训练**是 RBM 的一个重要算法，包含了正向过程、反向过程和权值更新 3 个步骤，主要目标是使生成的数据与原数据尽可能相似，

并通过对比两者的差异来调整权值更新：

$$w(t+1) = w(t) + a(vh^T - v'h'^T)$$

其中， a 是学习速率。这样的网络具备感知对输入数据表达程度的能力，而且尝试通过这个感知能力重建数据。如果重建出来的数据与原数据差异很大，那么进行调整并再次重建。

2012 年 6 月，《纽约时报》披露了 Google Brain 项目，吸引了公众的广泛关注。这个项目是由著名的斯坦福大学的机器学习教授 Andrew Ng 和在大规模计算机系统方面的世界顶尖专家 Jeff Dean 共同主导，用 16 000 个 CPU Core 的并行计算平台去训练含有 10 亿个节点的深度神经网络（DNN，Deep Neural Networks），使其能够自我训练，对 2 万个不同物体的 1 400 万张图片进行辨识。在开始分析数据前，并不需要向系统手工输入任何诸如“脸、肢体、猫的长相是什么样子”这类特征。Jeff Dean 说：“我们在训练的时候从来不会告诉机器：‘这是一只猫’（即无标注样本）。系统其实是自己发明或领悟了‘猫’的概念。”

2014 年 3 月，同样也是基于深度学习方法，Facebook 的 DeepFace 项目使得人脸识别技术的识别率已经达到了 97.25%，只比人类识别 97.5% 的正确率略低那么一点点，准确率几乎可媲美人类。该项目利用了 9 层的神经网络来获得脸部表征，神经网络处理的参数高达 1.2 亿。

最后我们再回到大数据这个时代背景上来。当坐拥海量的大数据，我们无论是做推荐系统还是 3D 模型检索（见第 6 章的 6.4 节“众里寻她千百度——海量 3D 模型的检索”），以前用简单的线性数学模型，一般也能获得还不错的结果。因此我们沾沾自喜起来，认为还是大数据更重要，而智能算法用简单直接的就 OK 了，不需要也没必要弄得很复杂。而当深度学习出现后，它的一系列辉煌战绩让我们意识到：也许是时候该“鸟枪换炮”了。简而言之，**在大数据情况下，也许只有比较复杂的模型，或者说表达能力强的模型，才能充分发掘海量数据中蕴藏的价值信息**。更重要的是，深度学习可以自动学习特征，而不必像以前那样还要请专家手工构造特征，极大地推进了智能自动化。

深度学习（即所谓“深度”）应**大数据**（即所谓“广度”）而生，给大数据提供了一个深度思考的大脑，而**3D 打印**（即所谓“力度”）给了智能数字化一个强健的躯体，三者共同引发了“大数据+深度模型+3D 打印”浪潮的来临。

第5章

3D智能数字化与3D照相馆：科学与艺术的结合

国学大师王国维在词牌《蝶恋花》中有云：“最是人间留不住，朱颜辞镜花辞树”。人逐渐老去，照镜子时已找不到年轻时候的朱颜；花谢了，纷纷从树枝上掉落下来。虽说这都是人世间不可避免的自然规律，但岁月蹉跎催人老着实令人怅然和无奈！

那么，我们真的留不住自己的“朱颜”么？即便阻止不了时光的流逝，把自己刹那芳华的某个瞬间留下也好啊，于是从古代的画像、制作面膜，到近现代的摄像拍照，智慧的人类不断地发明新的方法。在 3D 扫描和 3D 打印出现之前，制作蜡像是最逼真的三维人像保存方法。杜莎夫人蜡像馆是全世界水平最高的蜡像馆之一，蜡像馆是由蜡制雕塑家杜莎夫人建立的，有众多世界名人的蜡像，蜡像经常令人真假难分。

然而对于普通的民众而言，期待的不是一个只拥有世界级名人的蜡像馆，而是一个能以我们自己和家人为主角的人像陈列馆，我们也同样能将青春或者记忆用这样的形式珍藏，我们同样可以把自己摆放，我们也同样可以为世界留下些浮光片影。

有了 3D 智能数字化和 3D 打印技术，我们可以将自己和家人的 3D 模型自由摆放和组合，创造出各种让人回味的场景联想。如图 5-1 所示，想象一下：小时候 / 依稀 / 印象中 / 的 / 爷爷 / 与 / 现在长大 / 的 / 你，肩 / 并 / 肩 / 微笑着 /，一起 / 抬头 / 仰望天空 / 的感觉？



图 5-1 3D 打印：打造以个人家庭为主角的人像陈列馆（图片来源：creativevisualart）

5.1 那些年，我们一起追过的3D照相馆

我们先给 3D 照相馆一个定义。在传统的照相馆中，摄影师使用单反相机给客户拍照，然后将 2D 数码照片打印或者彩扩冲印出来。而在 3D 照相馆中，操作人员使用 3D 扫描仪对客户进行扫描，然后将美化修补后的 3D 数字化模型用 3D 打印机打印出来。

5.1.1 细数国内外的 3D 照相馆

目前，3D 照相馆已成为各大媒体上的热点词汇，下面我们就一起来追溯一下出现过的 3D 照相馆。

全世界第一家 3D 照相馆——西班牙 ThreeDee-You

特点：作品制作精良，经营时间久，已经在欧洲服务了很多客户，口碑很好。

在西班牙首都马德里，有家叫 ThreeDee-You 的摄影工作室早在 2010 年 6 月 27 日就已开始为人们制作 3D 打印的立体雕塑，如图 5-2 所示。在这家照相馆里，3D 打印的顾客被要求先进行三维扫描（如采用 OpticScan 三维扫描仪），获取全身的三维数据。顾客只需要保持姿势 2.5s 即可。接着，顾客会挑选自己喜欢的配饰或背景，然后通过软件进行调整。他们可以自由地选择 pose（造型），如蹲下、站立、微笑、打手势，甚至想要表达的个性化理念、形状、颜色。并且，这个过程可以重复多次，直到客户满意为止。

3D 人像价格从 94.50 欧元（约合 782 元人民币）到 289.50 欧元（约合 2 396 元人民币）不等，取决于人像的大小。



图 5-2 全世界第一家 3D 照相馆——西班牙 ThreeDee-You

全世界第二家 3D 照相馆——迪拜 Precise

这家照相馆在商场预约，制作时间较长。一个 6 inch（15.24 cm）的全彩雕像售价 300 美元（约

合 1 836 元人民币) 到 500 美元 (约合 3 060 元人民币) 不等。这种雕像使用塑料树脂, 全彩打印, 能包含真实的照片纹理。

全世界第三家 3D 照相馆——日本 Omote

日本商人的营销技巧一直不错, 网络上铺天盖地的报道都在热炒 Omote 是世界上第一家 3D 照相馆, 但这却并非事实。3D 照相馆 Omote 3D 在日本开张, 如图 5-3 所示, 打印价格如下。

- 小尺寸 (最大 10 cm, 20 g), 264 美元 (约合 1 616 人民币)。
- 中尺寸 (最大 15 cm, 50 g), 403 美元 (约合 2 466 人民币)。
- 大尺寸 (最大 20 cm, 200 g), 528 美元 (约合 3 231 人民币)。



图 5-3 全世界第三家 3D 照相馆——日本 Omote

下面再介绍一下国内的 3D 照相馆。

1. 国内第一家 3D 照相馆——西安非凡士

采用 MakerBot R2 打印机, 打印单色 3D 人像。公司的主营业务是机器人和机器狗。

2. 国内第二家 3D 照相馆——北京上拓 3D 打印照相馆

为京城首家, 上拓科技旗下电子商务平台“叁迪网”的线下体验店。打印全彩 3D 人像。

3. 国内第三家 3D 照相馆——武汉 3D 记梦馆

品牌经营, 连锁加盟, 面向全国市场。全彩 3D 打印人像, 10cm 高的人像售价 1 500 元左右, 15cm 与 20cm 高的则分别超过 2 000 元与 3 500 元。

5.1.2 3D 照相馆的设备及成本

传统的平面摄影是单反相机配合摄影棚完成的，而人像的三维数据采集则要借助一些更高科技的人像三维扫描设备。

3D 扫描设备及成本

目前，市售的手持扫描仪最低仅需 3 万元，贵些的在 10 万元到数百万元不等。比如采用国外进口的 Artec 3D 手持扫描仪（约 15 万元左右），在旋转平台上保持 5 ~ 15 分钟不动后，个人 3D 数据就可以出现在计算机的屏幕上了。此外，Kinect 扫描仪是一个廉价的考虑（一般为几千元），但扫描精度低，可用于个人试玩。

除了 3D 扫描，还可通过拍摄多张不同视角的照片来进行 3D 重建。目前，实现 3D 照片建模的软件不少，123D Catch 是著名的 Autodesk 公司开发的云端 2D 照片转 3D 软件，国内也有一款类似的软件叫 3D Cloud，但它们最大的缺点是需要云端上传，容易出问题。

3D 打印设备及成本

目前，国内的 3D 人像打印，性价比较高的是采用个人 3D 打印机，价格从 1 万元到 5 万元不等。但缺点是单色打印，如果客户要求，可能需要进行后期的表面上色处理，对美术功底有较高要求。单色 ABS 或 PLA 打印材料一般每千克成本约 200 元。

如果需要全彩打印，普遍采用 ProJet 系列 460（原 Zprinter 450）以上型号的工业级 3D 打印机，五六个墨盒调色后可以达到数十万种色彩，打印材料为高性能复合材料（也可以用石膏）。设备价格一般为 50 万元以上。塑料成像所用的高性能复合材料每千克成本约 3 000 ~ 4 000 元，石膏成像用的材料一般每千克成本约 500 元，打印时用到的黏结剂和后期用的后处理胶水都是按毫升计价，还要再加上各种颜色的墨水成本。因此建议将一批人像同时打印，这样可有效降低成本。



注意：在国内开设 3D 彩色照相馆，还需考虑到“野蛮快递”问题。采用石膏粉打印的 3D 彩色人像，强度不是很高，如果被快递员在运输过程中随意扔放，可能会发生损坏，特别是在打印的耳朵、手指、脖子等部位，而这将极大地影响到顾客的体验。

此外，3D 照相馆还有一个技术上的难点，即目前的 3D 扫描仪一般需花费几十秒甚至数分钟才能获取一个 3D 人像模型。相比之下，目前的 2D 照相机可以达到毫秒级的快门速度，以至可以捕捉到人脸表情瞬间微妙变化的神和魂，而这恰是照相馆中专业摄影师能够带给顾客的核心价值所在。因此，目前的 3D 照相离精致尚有一定的距离。

打印人像的所用时间是无法确定的，因为每个人像的大小不同，打印的时间也就不同。而许多想开办 3D 照相馆的人仍然想知道一个大概的时间，因此，下面将对以下例子进行说明，仅供参考，具体时间应以实际为准。

- 3D 全彩人像：3 ~ 5 小时（参考模型高：110mm，横向宽度：100mm）
- 单色人像：4 ~ 7 小时（参考模型高：110mm）
- 3D 水晶内雕人像：5 分钟（参考模型高：160mm，宽：120mm，长：80mm）

其他类型的 3D 成像业务

在上面的介绍中，出现了 3D 水晶内雕，那么我们就一并介绍一下其他的 3D 成像业务。

如图 5-4 所示，这是三维激光内雕机的成像效果。以 Spark 品牌为例，低成本精致级激光内雕机的价格一般为 20 万元左右。



图 5-4 三维激光内雕机的成像效果（图片来源：先临三维）

如果在水晶的底座上布置上七彩灯，则可以呈现出五彩缤纷的效果，如图 5-5 所示。



图 5-5 在水晶的底座上布置上七彩灯

此外，针对医院的大量客户群体，还可推出 3D 打印胎儿脸部模型的服务，如图 5-6 所示。准妈妈只需要提供一张超声（Ultrasonic，如三维 B 超）扫描图像即可。每个 3D 模型售价约 3 000 ~ 5 000 元。你可以选择 1:1 大小的模型，也可以选择适合随身携带的“迷你版”，只有普通手机链大小。大部分客户选择在怀孕八九个月时制作宝宝脸部模型。



图 5-6 3D 打印胎儿脸部模型（图片来源：Dos Santos）

日本东京一家名为 FASOTEC 的医学工程公司曾为客户提供 3D 打印技术制作的完整胚胎模型,胚胎模型由与子宫形状一致的透明材料包裹(如图 5-7 左边所示),宽约 9cm,取名“天使之形”,售价 10 万日元(约合人民币 6 142 元)。不过,制作“天使之形”要求准妈妈必须提供磁共振图像(MRI,Magnetic Resonance Imaging)或 CT(Computed Tomography,X 射线计算机断层成像)图像,鉴于孕期进行磁共振扫描存在一定的危险性,FASOTEC 已经限制了这项业务。而如图 5-7 右边所示的是温州一家妇产医院利用四维彩超和 3D 技术打印胎儿的立体模型。

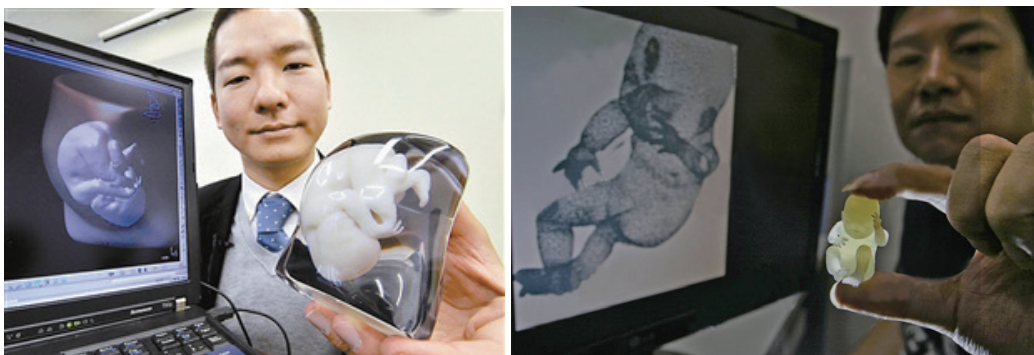


图 5-7 基于磁共振图像制作的完整胚胎模型(图片来源:FASOTEC、温州物像馆)

在打印材质的选择上,除了前面提到的 ABS、PLA、高性能复合材料、石膏等,还可使用金属材料,如金、银、铜、铁,如图 5-8 所示。工艺上一般采用铸模浇灌或者用 3D 雕刻机雕刻。



图 5-8 3D 打印纯银人像雕塑(图片来源:3D 记梦馆)

5.1.3 3D 照相馆赢利模式的探讨

中国目前有 13 多亿人口,每年登记结婚的新人超过 600 万对,再加上年轻人的写真照、儿童照、情侣照、孕妇照、老年人金婚照、全家福照等,这是一个年产值上千亿元的庞大市场。

3D 照相馆就是其中一个极具吸引力的商业模式,结合先进的 3D 扫描和 3D 打印技术,定做一个和真人同样姿势、状态、表情的微缩雕塑。我们回顾影像行业的发展,从黑白到彩色,现在是从 2D 到 3D。传统的 2D 摄影要立足于行业,无非是在摄影质量与装饰上下工夫,已很难有所突破,而 3D 打印人像则可能成为影楼差异化竞争的最佳武器。随着 3D 人像照相馆在全球范

围遍地开花，无疑会给婚庆和记录方式带来新一轮革命，也为更多艺术创作带来了全新的表现形式。

针对目前媒体对3D照相馆的密集报道，社会上也出现了不同的声音。杭州讯点科技CEO茹方军在微博点评称：“有点过热了”。他提到：“购买一台Zprinter 650售价60万元左右，人像扫描仪20万元左右，再加上场地费用及聘请专业工程师的开支，第一年大概需要150万元左右。在实际运作中，场地的日常维护需要费用，耗材大概每千克几千元，如果扣除税收等支出，利润其实并没有想象的高。”而打印一个人像，还有时间成本，每个人像大概需要打印3小时，“那么，在假设完全不存在失败率的情况下，每天一台机器持续运作，可以打印8个人像。这导致收回成本的周期变得相当漫长。”

“还得考虑到设备与技术更新的周期。”茹方军进一步质疑，目前3DP技术表面颗粒感还很强，但技术更新很快，现有的3D照相馆如何面对这种更新换代也会是一个问题。

然而，茹方军提到的一些具体数据其实是有待商榷、并不全面的。实际上，上面的设备报价都是针对国外进口产品的，价格当然非常昂贵。目前，“中国智造”的优质模式已经开始形成，并将逐渐席卷全球。以中国科学院的“视科三维”科技所研制的3D扫描和3D打印设备为例(<http://www.sik3d.com>)，给出了3D照相馆的全套解决方案，可将设备总成本控制在5万元以内。其中，所研制的手持式人像扫描仪，价格不到3万元，却比国外的同类产品（如Artec 3D）速度更快（每秒可达30帧以上）、效果更好；国产的3D打印机造价也不到1万元，精度却比MakerBot更高、速度更快。

在赢利模式上，3D照相馆其实可以有3个主营方向。一是开发刚结婚的新人和家庭客户，比如给准备结婚的新人打印婚纱人像。二是利用3D打印机去服务工业领域的产品样件制作、模具制造、小批量生产，比如打印玩具样品、个性化礼品/纪念品定制。当然，也可利用3D扫描仪提供工件扫描服务。三是作为代理商为国内外公司代销3D扫描/打印设备。

如果资金实在有限，希望尽可能地降低设备和人员成本，还可考虑智能云网模式（参见第4章4.5节“智能云网：云端智能服务和云制造”）。在这种新模式下，3D照相馆只需要购置一台3D扫描仪，然后将扫描得到的原始3D数据上传给云端的服务系统即可。云端服务系统快速进行3D模型的修补和美化，并自动选择距离客户最近的云制造节点将3D模型打印出来，并立即发货。在这个过程中，3D照相馆既无须掌握专业的3D数字化处理技术，也无须购买和维护专门的3D打印设备，相当于把这部分工作都外包给了云端的服务系统，极大地降低了资金和技术门槛。

5.2 3D照相馆的核心技术：3D智能数字化

开过照相馆或婚纱影楼的读者可能会有体会，照相馆的核心技术在于摄影技术，即要为客户拍摄出满意的照片效果来，然后将修正了的图像文件提供给工厂打印或冲印。同理，3D照相馆的核心环节也是对客户3D人像数据进行采集和处理，也即3D智能数字化技术。下面，我们就详细介绍一下3D人像数据的采集和处理过程，以便让你对各个环节都有一个直观的认识。

这里，我们以一款国外的 3D 扫描仪 Artec Eva 为例进行介绍。如图 5-9 所示，Artec Eva 手持式 3D 扫描仪酷似一台配有 3D 捕获功能的摄影机。该扫描仪无须标定，最高的捕获精度可达 16 帧每秒，帧图像可自动拼接对齐。Artec Eva 3D 扫描仪有 3 个摄像头，其中中间的带有一圈 LED 进行照明，主要用于获取颜色，另外两个摄像头能够得到与被拍摄者的距离数据，生成 3D 模型。Artec Eva 3D 三维扫描系统基于结构光技术原理（详见第 4 章 4.3.1 节“光学三维扫描仪的原理和实例（激光，结构白光）”），将特殊的光带，以成一个视差角的方式投射在物体表面。利用物体表面对光源所造成扭曲的原理精确计算每个三维数据点的坐标。

下面我们介绍详细的扫描步骤。

1. 按下按钮。对准扫描对象并按下按钮，如图 5-10 所示，扫描过程就会立即开始。操作非常简便。



图 5-9 Artec Eva 手持式 3D 扫描仪



图 5-10 按下按钮

2. 移动扫描仪。绕着扫描对象移动扫描仪，如图 5-11 所示。实时的表面对齐可以使你很好地了解已扫描了哪些部分，还有哪些部分没有扫描。如果你在某一个区域中无法获取扫描形状，请不要着急，稍后还可以返回去再扫描。

请根据需要尽可能多地扫描捕获完整的对象。如果需要旋转扫描对象以获取各个角度的扫描形状，请先完整地扫描一侧，然后关闭扫描仪，将扫描对象转至另一侧再对其进行扫描。

3. 将扫描形状对齐。将所有扫描片段对齐在一起后可以得到完整的模型，如图 5-12 所示。如果某些位置缺失，可以对此部分重新扫描一次。通过 3D 智能数字化算法，可以将多个扫描形状片段完美地对齐在一起，放置在一个统一的 3D 坐标系中。



图 5-11 移动扫描仪

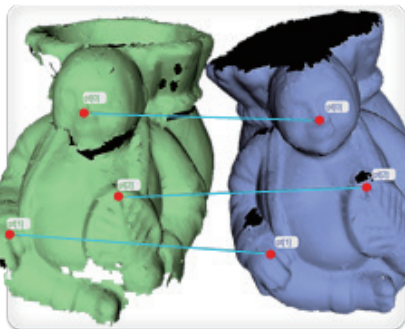


图 5-12 扫描形状片段的对齐

4. 将扫描形状片段融合成一个 3D 模型。将所有的扫描形状片段融合在一起，将会得到一个单一的三角形网格模型，如图 5-13 所示。



提示：每个扫描形状片段通常是由大量三维点组成的**点云（Point Cloud）**数据，将它们融合后得到的全局模型一开始也是点云数据。为了将点云模型转变成带有拓扑连接的三角形网格模型，就需要对点云进行**重建（Reconstruction）**或称之为**封装**。常用的重建方法有**泊松（Poisson）**重建、构造**等值面（Iso-Surface）**的**Marching Cube（MC）**算法、**德劳内（Delaunay）三角剖分（Triangulation）**（如 α -Shape、Crust、Cocone、Power Crust 算法）、基于区域增长算法（如 DBRA、BPA 算法）等。

5. 对扫描物体表面进行光顺和优化处理，如图 5-14 所示。还可以优化网格，填补孔洞并进行表面光滑处理。



图 5-13 将扫描形状片段融合成一个 3D 模型

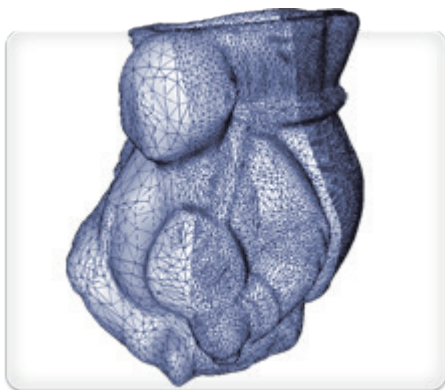


图 5-14 对扫描物体表面进行光顺和优化处理

6. 纹理图像贴图，如图 5-15 所示。轻敲一下鼠标键就可以自动地将纹理图像应用到扫描对象上。



图 5-15 纹理图像贴图

7. 得到原始的 3D 模型后，一般不能直接进行 3D 打印，还需要做一些后期处理，对模型进一步优化，这一点和我们用 Photoshop、Lightroom 处理照片很像。具体来说，扫描得到的原始 3D 数据一般都含有噪声，而且头发的效果往往很差，普遍采用的是 Geomagic 配合 ZBrush

或 3D-Coat 进行人像处理。下面，我们就将 3D 照相馆需要用到的 3D 智能数字化技术一一罗列出来，供读者参考。

- Geomagic 具有三维扫描数据处理功能，可对缺失和噪声数据进行修补，并拼接得到一个完整的 3D 模型。**具体请参见 5.5.3 节“Geomagic Studio：更通用的任意形状修补”。**
- ZBrush 或 3D-Coat 是三维数字雕塑软件，非常适用于头发的修补和人脸的后期打磨。**具体请参见 5.5.1 节“使用 3D-Coat/ ZBrush 软件手工修补发型”。**
- 另外，Magics 也是 3D 文件打印前理想的软件解决方案，它能够快速高效地修正具有瑕疵的打印文件。**具体请参见 5.8.2 节“Netfabb/Magics：修正你的 STL 打印文件”。**
- 如果顾客对扫描得到的人像不满意，比如说觉得不够漂亮，可对 3D 人脸形状和图像纹理进行美化。**具体请参见 5.3.3 节“人是种视觉动物：如何美化你的照片”。**
- 还可以对 3D 形状进行一些编辑和修改，比如顾客的背有点驼，希望能够变成笔直的；或者觉得自己的鼻梁有点塌，希望高挺一些；甚至很多女性顾客希望自己的身材更加凹凸有致；还有的男性顾客希望自己的手腕上能多上一块劳力士手表，甚至希望给自己穿上钢铁侠的装甲等。**具体请参见 6.2.3 节“个性化形状的编辑与合成”。**
- 相比于 2D 照相，3D 照相的另一个巨大优势是有更大的操作自由度。比如顾客在照相时是一种正常的表情，在打印输出时，我们可以对 3D 表情进行编辑，改为微笑或大笑的表情。**具体请参见 5.6 节“3D 人脸表情形变与编辑”。**
- 使用 3D 扫描仪的主要缺点是顾客需要保持一个姿势至少几分钟，成人还可以忍受，但对于活泼好动的小朋友或小猫、小狗来说，它们可能连 1s 都无法坚持。这时，就需要用到基于多张照片的 3D 重建技术了，可在影棚中的各个方位架设多台（比如 20 台）照相机，在毫秒级的时间内同步拍摄。**具体请参见 5.3.2 节“基于多视角照片的 3D 人脸重建”。**
- 有时候顾客会摆出一些很奇怪的姿势，比如金鸡独立、街舞倒立或腾空而起等，那么如何让这些奇怪姿势的 3D 模型打印出来能够放稳呢？**具体请参见第 6 章 6.7 节“形状平衡：如何确保 3D 物件站立稳当”。**
- 由于目前全国各地的 3D 照相馆还不是很多，所以经常会有不方便前来的外地顾客。这时可以让顾客提供一张清晰的正面照片，然后使用 3D 智能数字化技术重建出对方的 3D 模型。**具体请参见 5.3.1 节“基于单张照片的 3D 人脸重建及立体浮雕”。**
- 顾客总是希望能花更少的钱，这时店家可以从尽可能减少打印耗材着手削减成本。**具体请参见第 6 章 6.8 节“形状优化：生成坚固的内部轻质结构使得耗材最省”。**

5.3 基于图像的3D人脸重建技术

在 3D 照相馆中，除了 3D 扫描仪这种主动式扫描设备，还可采用被动式重建技术。这种技术对被测对象不发射任何光，而是通过采集被测物表面对环境光线的反射来获取数据，因此不需要规格特殊的硬件，往往只需要一台或几台照相机获取多个视角的图片即可，因此成本非常低廉。

比如外地的顾客不方便前来扫描，这时可以让顾客用普通相机拍摄一张清晰的正面照片并远程发送过来，然后利用 3D 智能数字化技术就可以重建出对方的 3D 模型。又比如，手持式 3D 扫描仪的扫描时间较长，顾客需要保持一个姿势至少几分钟。这时，如果在影棚的各个方位架

设多台照相机，在毫秒级的时间内获取到多个视角的多张同步照片，就可以利用立体视觉重建技术来生成瞬间的 3D 模型了。下面我们分别进行介绍。

5.3.1 基于单张照片的 3D 人脸重建及立体浮雕

3D 人脸建模技术可分为线性混合人脸、基于参数化曲面的人脸建模和设计、基于三维数据点插值的人脸重建、基于物理原理或生理仿真的人脸重建等。这里我们介绍一下目前广泛应用的线性混合人脸方法。

“形状混合人脸”技术认为所有人脸构成了一个线性空间。通过对配准对齐后的大量三维人脸模型及其纹理的主成分分析（PCA）降维技术，任意一个人脸均可以用有限个人脸基的线性组合逼近（详见第 6 章 6.2.2 节“个性特征的定位与匹配”）。这种方法要先建立一个具有类型多样性的三维人脸数据库。然后，通过主成分分析方法对人脸数据库中的众多人脸模型进行统计分析，提炼出可用于线性组合的人脸基，以此建立统计意义上的参数化通用人脸模型。

例如，德国研究人员 Blanz 和 Vetter 通过建立配准的西方人脸数据库，提出了称为“3D 形变模型”（3DMM, 3D Morphable Model）的参数化 3D 人脸模型。每张人脸网格的所有 N 个顶点坐标及其纹理颜色可组合成如下的形状向量 \mathbf{s} （含 X 、 Y 、 Z 这 3 个坐标分量）和纹理向量 \mathbf{t} （含 R 、 G 、 B 这 3 个颜色分量）：

$$\mathbf{s} = (x_1, y_1, z_1, \dots, x_N, y_N, z_N)^T \in \mathbf{R}^{3N}$$

$$\mathbf{t} = (r_1, g_1, b_1, \dots, r_N, g_N, b_N)^T \in \mathbf{R}^{3N}$$

通过对数据库中所有人脸的形状向量空间和纹理向量空间进行主成分分析，就可以将任何一张新脸的形状向量 \mathbf{s}^* 表示为平均形状向量 $\bar{\mathbf{s}}$ 和 M 个形状主元向量（即形状人脸基） $\boldsymbol{\varphi}_m$ 的线性组合；任何一个新人脸的纹理向量 \mathbf{t}^* 表示为平均纹理向量 $\bar{\mathbf{t}}$ 和 M 个纹理主元向量（即纹理人脸基） $\boldsymbol{\theta}_m$ 的线性组合，其中 α_m 、 β_m 为线性组合系数。

$$\mathbf{s}^* = \bar{\mathbf{s}} + \sum_{m=1}^M \alpha_m \boldsymbol{\varphi}_m$$

$$\mathbf{t}^* = \bar{\mathbf{t}} + \sum_{m=1}^M \beta_m \boldsymbol{\theta}_m$$

Blanz 和 Vetter 将该参数化人脸形变模型拟合到一幅正面图像上，重建该图像对应的三维人脸模型及其纹理，生成了较高真实感的人脸。该技术已经被广泛地集成于各种人脸建模和动画软件如 FaceGen、Poser、Maya 中，用于快速地、低成本地生成三维人脸模型，本节介绍的人脸建模就是采用了这种方法。如果读者对上面的阐述不太明白，强烈建议阅读一下第 6 章 6.2.2 节“个性特征的定位与匹配”，其实是非常简单、易懂的。

下面我们就来介绍如何使用 FaceGen Modeller 软件将单张正面照片生成三维人脸。实际上，它允许用户提供 1 ~ 3 张照片，其中正面照片是必须的。之所以可以提供额外的两张照片，是由于从单张正面照片生成人脸是一个病态（ill-posed, ill-conditioned）问题（即不适定问题，可

简单地理解为不确定问题)，所以有时在没有两个左右侧脸照片的情况下，一些侧面轮廓特征，比如鼻梁的高低是无法重现的。这其实很好理解，有时我们经常看一位美女的侧脸很美，紧紧跟了上去，结果对方猛一回头终于看见她的正脸了，却发现正脸和侧脸并不太搭配。

按照软件提示，第一步载入一张人脸正面照片，如图 5-16 所示，由于三维模型的纹理贴图用的是载入的照片，所以对照片的要求应该没有遮挡，光照均匀的，否则眼镜等物品将会直接被当作脸部纹理映射到模型上。



图 5-16 第一步载入人脸正面图像

接着我们根据向导提示，在人脸上手工标记出特征点，如眼角、鼻尖、嘴角等，如图 5-17 左边所示，以方便计算机进行特征定位。建模过程是一个迭代计算误差函数的过程，因此形状会不断演化，越来越接近照片中的样子，如图 5-17 右边所示。

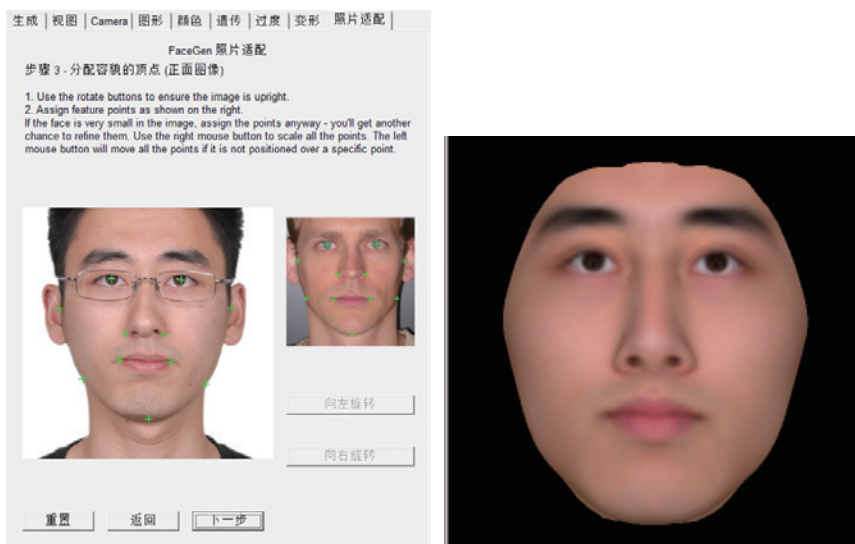


图 5-17 左：手工标记特征点；右：逐渐迭代逼近照片中的人脸

生成后的人脸模型可以导出为 obj、3ds 等多种不同的 3D 文件格式。还会同时导出各个部件（如眼球、嘴巴、牙齿）的纹理贴图。

下面是根据单张照片的 3D 人脸重建效果，如图 5-18 所示。

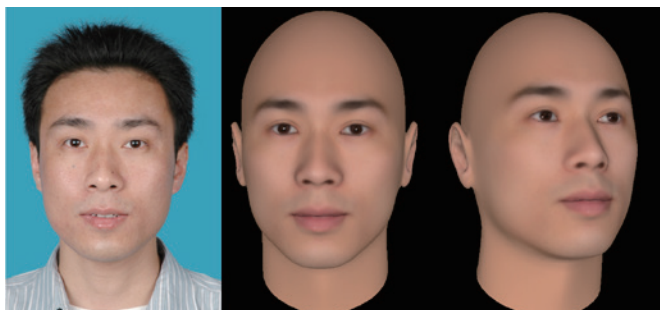


图 5-18 基于单张照片的 3D 人脸重建技术（图片为笔者本人）

此外，我们并不总是需要进行 360° 各个方位的 3D 重建，很多情况下只是希望正面能有个立体的效果即可，也即浮雕效果。因此，可利用**视觉计算**技术，我们对照片中场景的深度进行估计，自动生成带有立体效果的深度图，然后将其 3D 打印出来即可，如图 5-19 所示。当然，图像的分辨率越高，制作出来的立体浮雕效果也就会越好。



图 5-19 基于单张照片的立体浮雕打印效果（图片来源：BumpyPhoto）

5.3.2 基于多视角照片的 3D 人脸重建

多视角三维重建的技术原理请详见第 6 章 6.3 节“立体视觉重建：将照片转成 3D 数字模型”，本节主要介绍如何具体地操作和应用。Autodesk（欧特克）公司发布了一个建模软件 Autodesk 123D Catch，有了它，你只需要简单地为物体拍摄多张照片，不需要复杂的专业知识。利用云计算的强大能力，123D Catch 可以将用户拍摄的照片迅速转换为逼真的 3D 模型。官方下载地址 <http://www.123dapp.com/catch>，下载安装注册都很容易。

在实际使用时，用户一般需要拍摄至少 20 张左右不同角度的照片。比如你可以让朋友对着你的头部拍摄一组照片，然后用 123D Catch 生成你头部的 3D 模型。如果可能，最好围绕着物体拍摄两圈，每圈照片错开角度、高度和间隔，如图 5-20 所示。



图 5-20 拍摄至少 20 张左右不同角度的照片（图片来源：Autodesk）

再举一个例子，如图 5-21 所示的这组照片是先以约 60° 的角度环拍一圈，再以约 15° 的角度环拍一圈。在拍摄之前，先计划好如何移动，如何从不同角度拍摄。记住一点，先围绕着物体转一圈拍摄，然后再拍些细节。围绕物体拍摄时，多少度角照一张，取决于你想拍的物体。有些物体 25° 角就可以。有些样子复杂的物体，需要更密集的拍照，比如转 10° 就需要照一张，转一圈下来共需要拍 36 张照片。

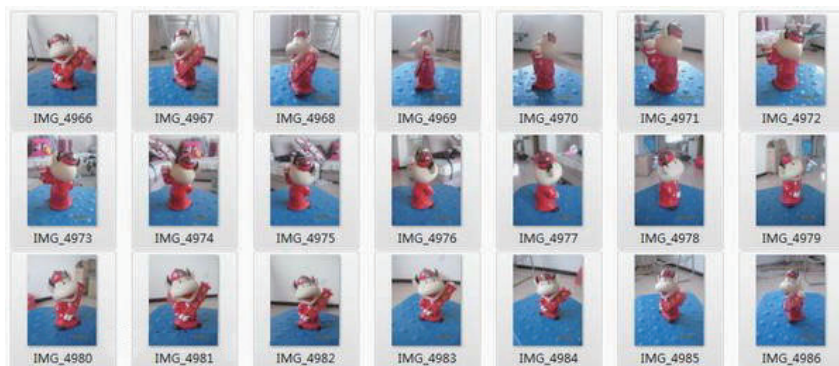
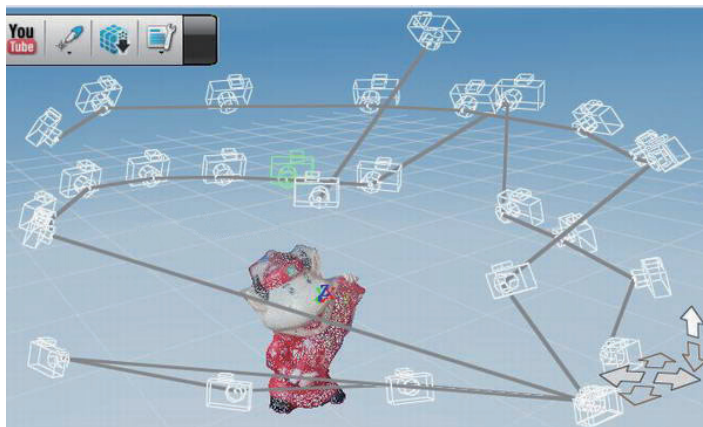


图 5-21 拍摄的路径规划以及所拍摄的多角度照片（图片来源：3dprintime）

根据 Autodesk 的官方说明,要想用这个软件做出理想的 3D 模型,你还必须掌握如何去拍照,需要注意的地方如下。

- 整个拍摄过程,要保持在同样的光照下,有专业的环境最好。光线不能太亮或太暗。拍摄时也不能用闪光灯。拍摄时还要注意相机不能抖动以免照片模糊。
- 要连续拍完,不能今天拍了几张,明天再拍几张。
- 避免拍摄透明、反光或平坦光滑的物体。
- 拍摄时,也要避免有对称特征的物体。
- 相邻照片之间最好有 50% 的场景重合。这样软件才能根据相同部分,把其他不同部分加进来。
- 如果拍摄活的动物或人,要确保他们在拍摄过程中不动。此外,不停的乱动也会使快门速度不够高的普通相机拍出模糊的运动图像来。
- 拍物品时,也不能把物品翻过来拍。把想拍的物体放好后,围着它转着拍。拍摄中只能你动,不能移动物体。

123D Catch 采用的是**多视角立体匹配重建 (MVS, Multi-View Stereo)** 技术,原理建立在以特征匹配为基础的三角测距上。因此,拍照时尽量把物体放置在有特征的背景里,否则软件很难在多张照片之间找到匹配,导致无法生成 3D 模型。比如,要避免拍摄没有特征的背景墙,软件非常难识别这样的背景照片。你可以添加一些特征到墙上,例如,贴一些画是个好办法。又比如,你想拍一只恐龙,不要把它放在光滑的地板上。可以找块有特征的地毯或铺一张报纸(如图 5-22 左边所示),把恐龙放在地毯/报纸上面再拍摄。当然,你也可以在背景里粘贴一些特征纸条,如图 5-22 右边所示。同理,也不能拍摄那些平行、相似度太高(特征模糊的)的物体,例如,有很多一模一样窗户的楼房。

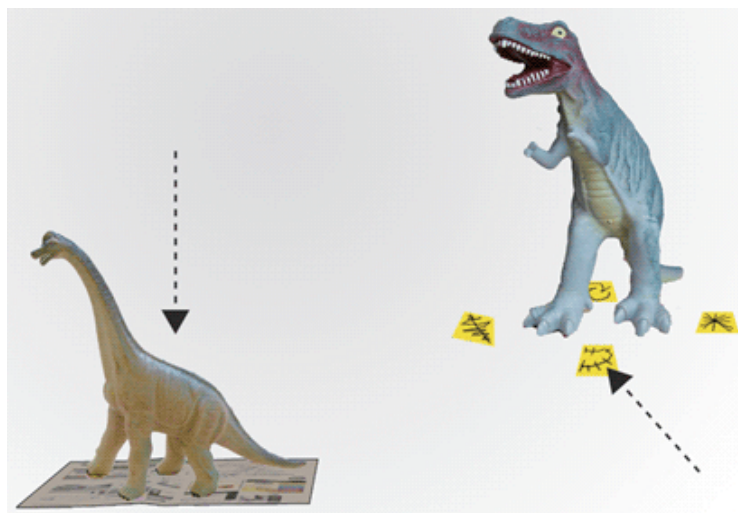


图 5-22 将物体放置在有特征的背景里。左：报纸；右：画有特征的纸条

拍完照片之后,你就可以上传了。注意在上传前,你最好不要修改原始照片。因为原始照片带有照相机的参数,软件需要这些参数进行标定。上传照片后,软件会自动找到和匹配照片中物体的共同特征。以这些共同特征为基点,把不同照片中物体的特征整合在一起,以便生成 3D

模型。如图 5-23 所示，这是所生成的 3D 模型和它的三角形网格。你可导出为 3D 文件，以便输出到 3D 打印机进行打印。

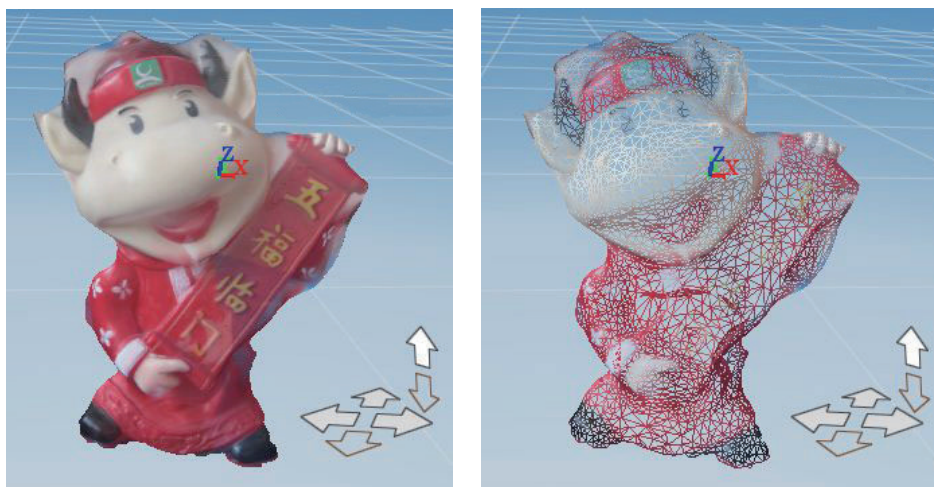


图 5-23 生成的 3D 模型和它的三角网格

通过阅读前面的拍摄注意事项，可能有的读者心里会犯嘀咕：拍照难道要这么讲究吗？非也！你当然也可以有不那么讲究的套路，但前提是需要加钱购置设备。如图 5-24 所示，我们可以绕顾客一圈上下架设多台照相机（比如 20 台），以毫秒级的精度控制它们同步拍摄。这样，你就不用费时费力地端着相机绕顾客拍了。



图 5-24 架设多台照相机同步拍摄人像（图片来源：NUS）

然后，使用 3D 智能数字化算法，比如利用多视角立体视觉、Visual Hull 技术（见第 6 章 6.3.2 节“基于立体视觉、SfM 和 Visual Hull 的三维重建”）以及几何细节形变技术（见第 6 章 6.2.3 节“个性化形状的编辑与合成”），我们就可以获得瞬间的 3D 形状。如图 5-25 所示，因为拍摄在毫秒级的时间内完成，这时顾客就不再需要保持一个姿势几分钟不动了。

除了同步拍摄人体，我们当然也可以同步拍摄人脸。如图 5-26 所示，我们可利用多台同步拍照的高清单反相机，基于多视角重建技术来获取“毛孔级精度”的精细 3D 重建^[69]。



图 5-25 使用 3D 智能数字化算法重建出瞬间的 3D 形状（图片来源：MIT）

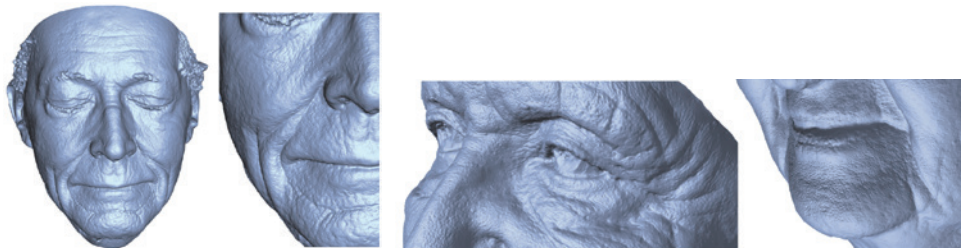


图 5-26 利用多台同步拍照的高清单反相机来获取“毛孔级精度”的精细 3D 人脸
（图片来源：ETH Zurich）

有的读者会说，上面的技术确实很酷，连毛孔都能拍出来。但这么多三脚架摆在顾客面前确实吓人（有种“大刑伺候”的感觉？），而且多台单反相机成本也不低，此外还有很多顾客脸上痘痘什么的也不少，毛孔级精度似乎太精细了（正所谓：架起大炮找蚊子）。没关系！创客们

最大的优点就是乐于成为普通大众的“贴心小棉袄”，比如牛津大学就研制出一款手持式 3D 扫描仪 Fuel3D，任何人都能通过它快速获得现实物体的 3D 模型。像普通相机一样，只需对准目标，按下快门，Fuel3D 就能在数秒内帮你抓取 3D 人脸模型。如图 5-27 所示 Fuel3D 左右安装有两个摄像头，也是基于计算机视觉立体成像的原理。

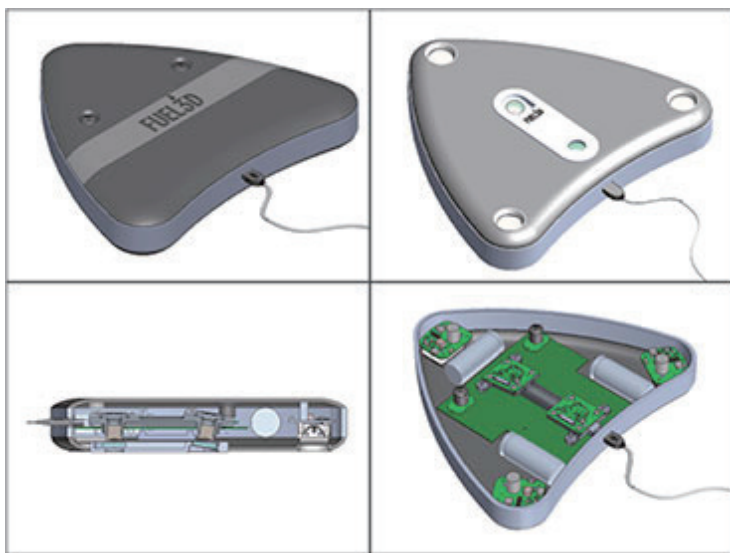


图 5-27 Fuel3D 的外形图（图片来源：牛津大学）

与 5.4 节的 Kinect 相比，Fuel3D 精度更高，如图 5-28 所示，这得益于它更高的拍摄分辨率。

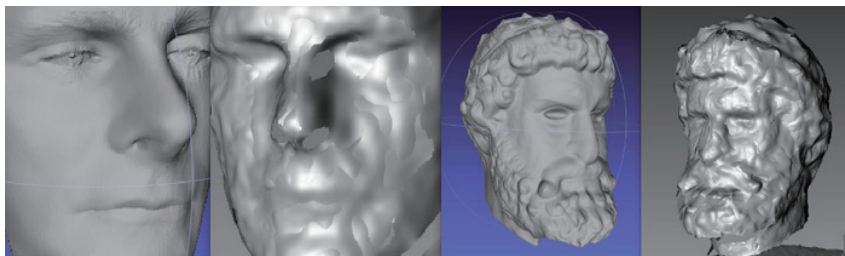


图 5-28 Fuel3D 与 Kinect 的扫描结果对比。每个小图的左边为 Fuel3D 的结果，右边为 Kinect 的结果。

5.3.3 人是种视觉动物：如何美化你的照片

所谓“爱美之心，人皆有之”。而随着视觉计算技术的发展，在一定程度上让计算机学习人脸“美丽”的概念成为了可能。

认知心理学近年来大量研究实验发现：人们对于什么脸是美的存在着高度的一致性认可，这种高度的一致与文化、种族、年龄、性别无关。同时，神经心理学和心理物理学利用磁共振成像等技术发现，美丽的面孔引起了大脑里奖赏回路（如杏仁核区域）的激活。换言之，当你看见美女或帅哥时，你的大脑会觉得非常受用，作为对你的奖励，大脑会让你的心情感到很开心和兴奋。

下面我们先用人女们熟悉的美图秀秀来演示一下,如何手工对人脸进行美化。如图 5-29 所示,通过美白肌肤、将眼睛扩大、向内拉伸脸部轮廓进行瘦脸,确实可以让女生变得更美丽。此外,更专业的美化软件有 Portrait Professional、Portrait+ 等。



图 5-29 使用美图秀秀进行人脸美化 (图片来源: 美图秀秀 xiuxiu.meitu.com)

美图秀秀虽好,但需要手工不断地去调整(也即不断地“修理”你的脸),而且很多缺乏美学细胞的男生根本就不知道如何去调。那么,能否有办法让计算机自动帮你美化呢?这涉及两个问题需要解决。首先,要让计算机知道什么样的人脸是美丽的,因为计算机不像人脑那样有奖赏回路(如杏仁核区域),所以它看见美女也不会怎么兴奋;其次,具体应该如何实现人脸美化,也即美化算法的实现。

好,我们首先告诉计算机:什么样的人脸是美丽的。近代的心理学研究表明,对许多张人脸图像进行平均合成,得到了一张合成人脸图像,其往往会比合成前的那些图像显得更美丽。这个理论最早在 19 世纪末由心理学家弗朗西斯·高尔顿(Francis Galton)提出。一般认为,平均脸是吸引人的,虽然最美丽的脸不一定是平均脸,但平均脸至少可以得到中等程度以上的吸引力,这就是心理学上著名的“**平均脸假说**”(Averageness Hypothesis)。如图 5-30 所示,给出了中国、日本、韩国、美国、印度、非洲这些国家和地区的男女平均脸。按笔者的观点,平均脸可以得到完美(Perfect)的脸,因为平均后不会存在什么缺陷(比如这里缺一块,那里鼓出一块来),虽然不一定是迷人的(Charming),但至少在一般人看来是美丽的(Beautiful)。平均脸的详细合成方法请参见第 6 章 6.2.2 节“个性特征的定位与匹配”。

除了平均脸模型,当然我们也可以对人脸数据库的每张人脸图像进行人工打分,通过对打分集的训练,来量化地告诉计算机什么样的人脸漂亮,并可以打几分。这是基于数据驱动(Data-driven)的人脸美化的技术思路。所谓的基于数据驱动的人脸美化算法,是要从一个巨大的人脸图像数据库中学习到人脸美化的准则,通过学习到的准则,我们可以对任意一张输入的人脸进行美化。

接着,我们来解决第二个问题。人脸美化算法的主要目的在于,在保留原始人脸的基本特征的前提下,使得美化后的人脸获得尽可能高的美学吸引力。其中有两个关键点,一是原始人脸的基本特征要得到尽可能地保留,即美化后的人脸看上去还是同一个人;二是获得尽可能高的美学吸引力,即尽可能地美化一张人脸。

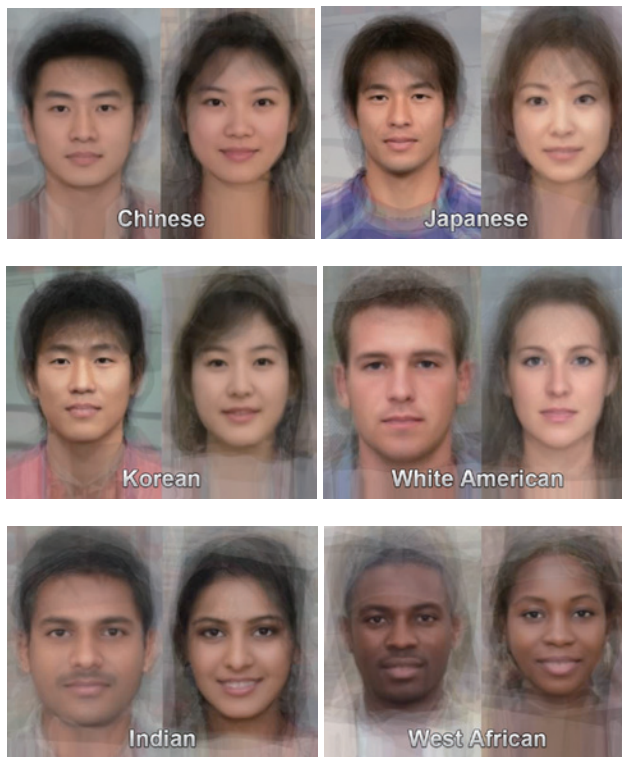


图 5-30 中国、日本、韩国、美国、印度、非洲这些国家和地区的男女平均脸（图片来源：dailymail）

这里我们介绍一下 Tommer Leyvand 等人提出的基于 **KNN** (K Nearest Neighbors, K 近邻法) 的美化算法^[40]，如图 5-31 所示，可提高正面人脸照片的美丽（或吸引力），而同时保持与原始照片相似。算法的基本思路为，从人脸数据库找出 K 张与输入人脸图像最相似的人脸，并计算这些人脸形状的加权平均（权值由相似人脸与输入人脸的相似度以及美丽程度决定），从而得到一张平均后的目标人脸。再将输入人脸的形状变形到那张平均人脸形状，便可得到一张美化后的人脸。这一算法简单有效，它同时也验证了心理学上的“平均脸假说”，即平均脸一般要比用于合成它的各张原始人脸漂亮。



图 5-31 人脸美化算法的结果对比。上：美化前；下：美化后（图片来源：Tommer Leyvand）



提示：KNN 在经典的模式识别应用中常作为一种分类算法，非常简单、直观：给定一个数据库（比如含有很多人和猩猩），对于某个要判定的样本（比如一个人），在数据库中找到与它距离最邻近的 K 个样本（比如 9 个人和 1 只猩猩），这 K 个样本的多数属于某个类（这里是人类），就把要判定的样本归于这个类。正所谓：“近朱者赤，近墨者黑。物以类聚，人以群分”。

在距离度量上，有**闵可夫斯基距离**（Minkowski Distance，也被称为 L_p 距离）：

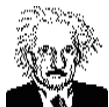
$$d^{AB} = \sqrt[p]{\sum_k |x_k^A - x_k^B|^p},$$

这个距离涵盖了：常见的**欧氏距离**（即 L_2 范数，对应于 L_p 距离中 $p=2$ 的情况，即两点间的直线距离）、**曼哈顿距离**（Manhattan Distance，即 L_1 范数，对应于 L_p 中 $p=1$ 的情况，即两个街区的行走而非直线穿墙距离。在 lasso 回归中可用 L_1 范数使得一些系数被迫缩减为 0）、**切比雪夫距离**（Chebyshev Distance，即**无穷范数**，对应于 L_p 距离中 $p=\infty$ 的情况，即从各个维度的单维距离中选取最大值）等。

然而，闵氏距离没有考虑到各个维度的**量纲**（Dimension）、**尺度**（Scale）和分布的不同，**马氏距离**（Mahalanobis Distance）则克服了这个问题，不仅与测量单位无关，而且表示了数据的协方差距离，排除了维度之间相关性的干扰。

在 K 近邻的快速计算上，有 kd 树搜索算法。

除了 KNN，还有一种基于 **SVR**（Support Vector Regression，**支持向量回归**）的美化算法，与 KNN 算法的区别在于，目标美化人脸的计算。基于 SVR 的人脸美化算法并不直接利用数据库中的 K 张相似人脸样本去计算目标形状，而是首先根据所有人脸样本和对应的打分拟合出一个评价人脸美丽度的函数 fb ，然后在待美化的人脸形状空间附近寻找一个使 fb 取得极大值的相似人脸形状，该形状便为目标美化形状。研究结果显示，无论是基于 KNN 的算法还是基于 SVR 的算法，通常都能够产生更美的人脸图像，这是非常有意义的成果。这也证实了人脸的美丽程度是可以被计算机所学习、量化和美化的。



提示：SVR 可认为是 SVM（支持向量机，将在 6.4.2 节介绍）用于解决**回归问题**（Regression）的一个扩展，本质上是相同的。经典的 SVM 一般用于解决**分类问题**（Classification）：判定一个样本所属的类别，以对其指定一个离散的整数标签（比如是人还是猩猩，属于正样本还是负样本，属于第 1 类还是第 2 类等）。而回归则需要给样本指定一个**连续的实数值**（比如这位女生的美丽程度可以打 8.63 分）。因此，SVM 用找到的超平面将所有样本一分为二，而 SVR 用找到的超平面去拟合连续的样本分布函数。

具体地，SVR 可写成如下的数学形式：

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & \|y_i - (w \cdot x_i + b)\| \leq \varepsilon \end{aligned}$$

其中参数 $\varepsilon \geq 0$ ，用来表示 SVR 预测值与实际值最大的差距。与 6.4.2 节和 10.1.3 节中介绍的 SVM 相对照，可发现十分相似，不同之处为 SVM 考虑的是预测类别 (Predicted Class) 和实际类别 (Actual Class) 需同号（即预测正确），而 SVR 考虑的是预测值和实际值的差距需小于 ε 。

目前在商业软件 Portrait Professional 中，如图 5-32 所示，已经提供了类似的自动美化功能，并允许用户对人脸的每个部位做进一步的美化调整。



图 5-32 Portrait Professional 软件的人脸美化功能

5.4 Skanect：使用Kinect实现3D扫描

Occipital 公司开发了一款 3D 扫描软件 Skanect，能够让用户使用廉价的视觉传感器（比如 Kinect、Asus Xtion Pro Live、Primesense Carmine 等），在 30s 内捕捉到室内和人物的全彩 3D 模型。

使用时，将 Kinect 连接到计算机上，打开 Skanect 软件。可以看到界面最上边有 Prepare（准备）、Record（录制）、Reconstruct（重建）、Process（后处理）、Share（分享）等步骤，如图 5-33 所示。

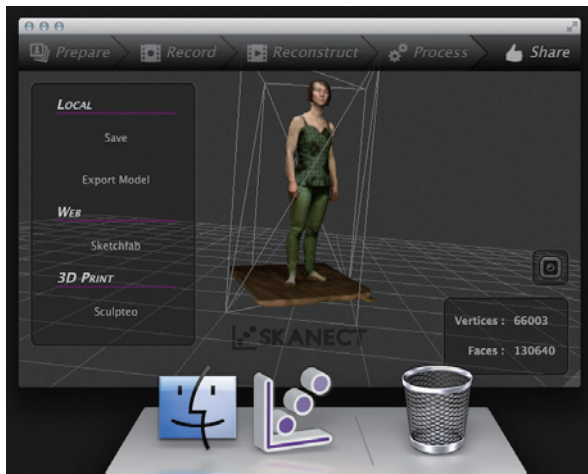


图 5-33 Skanect 主界面

在 Prepare（准备）阶段，我们选择记录的类型，人像要选 Body，而 BoundingBox 是指你要记录的人体大小，如果是半身像，1m×1m×1m 就可以了，单击“Start”按钮进入下一步骤。

现在我们可以开始录制了，如图 5-34 所示。注意：扫描的时候不要佩戴眼镜；还有最好戴帽子等，因为头发的扫描效果不好。单击倒计时按钮后经过 5s 倒数即可开始扫描。

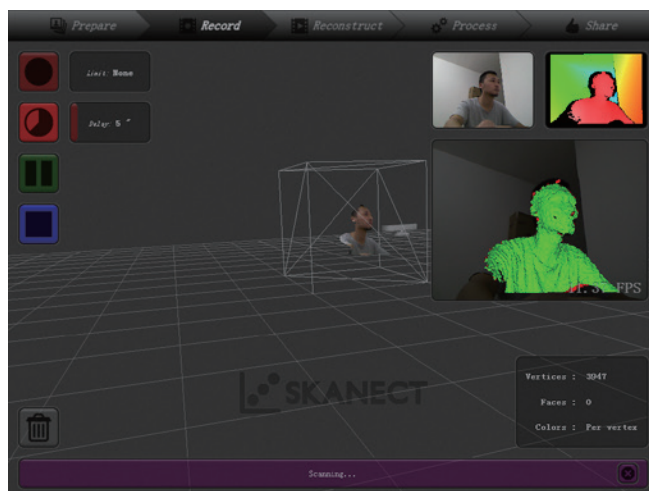


图 5-34 开始扫描形状

扫描的时候可以让朋友帮忙拿着 Kinect 围绕你转圈，或者自己在凳子上转圈都可以，如果有转盘的话效果会更好，扫描环境最好是在光线充足的地方，移动速度要缓慢而均匀。扫描完成后可得到一个半身的无色立体图像，如图 5-35 所示。

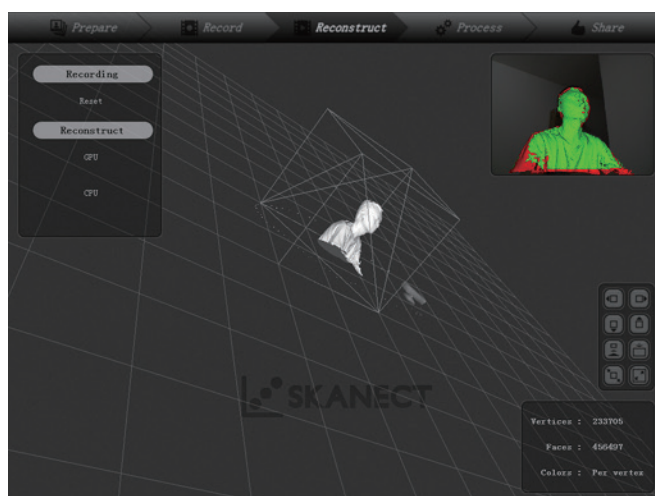


图 5-35 扫描后得到的半身无色立体图像

如图 5-36 所示，可以快速地处理一下数据，如填补孔洞、上色等。处理后可以导出 3D 模型，如选择导出带彩色纹理的 Ply 文件格式。

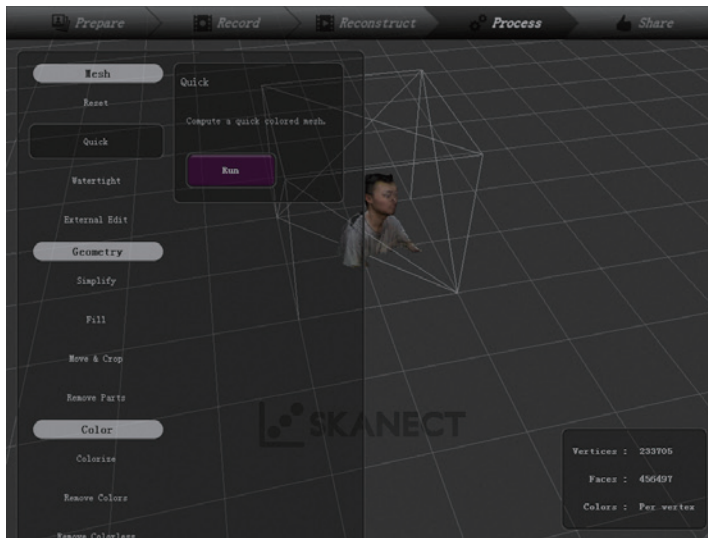


图 5-36 快速地处理一下数据，如填补孔洞、上色

然后用专业的网格处理软件 Geomagic 打开这个 Ply 文件，进行填孔、光滑、网格医生修复曲面等。具体请参考 5.5.3 节“Geomagic Studio：更通用的任意形状修补”。

除了 Skanect，还有两款类似的软件 SCENECT 和 ReconstructMe。SCENECT 由一家名为 FARO Technologies 的便携式 3D 扫描仪生产厂商免费提供，如图 5-37 所示，这是它的 3D 扫描效果。

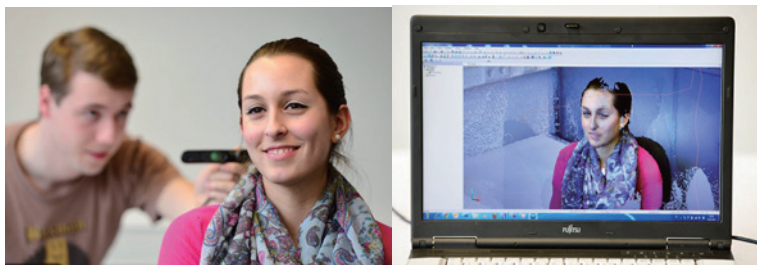


图 5-37 SCENECT 的 3D 扫描效果（图片来源：FARO）

另一款名为 ReconstructMe 的软件所扫描的 3D 模型和打印的效果如图 5-38 所示。

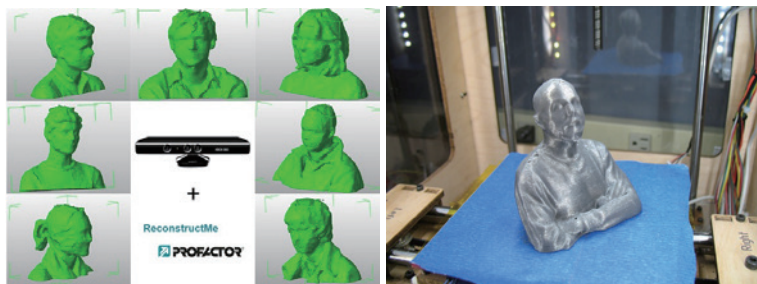


图 5-38 ReconstructMe 软件所扫描的 3D 模型和打印后的效果

5.5 头发修补：3D照相馆的头痛问题

与用数码相机拍出一张 2D 图像不同，原始的 3D 扫描数据往往有很多噪声和空洞，尤其是对于头发，扫描结果往往很差，如图 5-39 所示。因此，我们一般需要对原始的噪声和残缺数据进行修补。



图 5-39 原始的 3D 扫描数据往往有很多噪声和空洞（扫描对象为笔者本人）

下面我们将介绍如何修补发型，包括使用软件手工修补和基于视觉计算自动修补。我们还将重点介绍如何使用 Geomagic Studio 对任意形状进行拼接、修补、平滑。

5.5.1 使用 3D-Coat/ ZBrush 软件手工修补发型

目前有两款功能强大的 3D 雕塑软件 3D-Coat 和 ZBrush，都可轻松实现头发的修复。如图 5-40 所示，是使用 3D-Coat 软件对图 5-39 的残缺发型进行修复的效果。

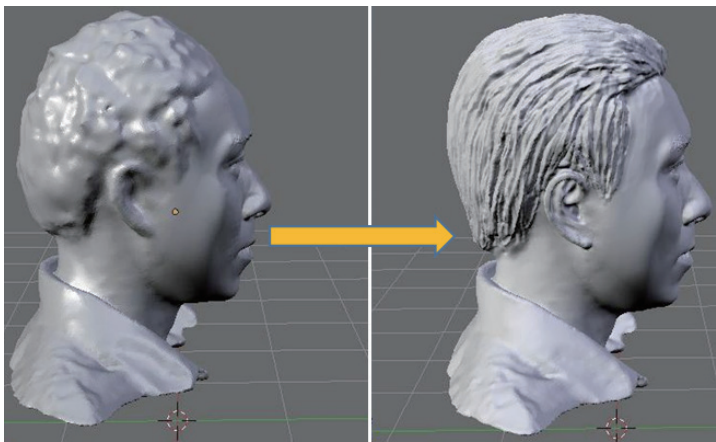


图 5-40 使用 3D-Coat 软件进行发型修复的效果（笔者本人）

下面对修复过程做一介绍。打开 3D-Coat 雕塑软件，选择“体素雕刻”模式，如图 5-41 所示。

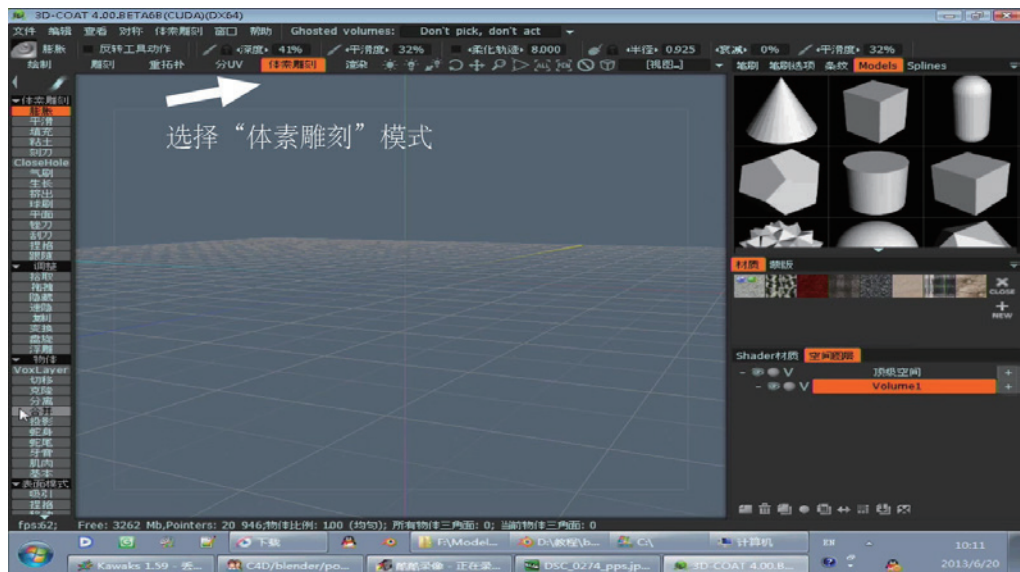


图 5-41 打开 3D-Coat 雕塑软件，选择“体素雕刻”模式

单击“合并”按钮，在弹出的子面板中单击“选择模型”按钮，选择需要编辑的原始扫描 3D 模型（如 obj 文件格式），如图 5-42 所示。

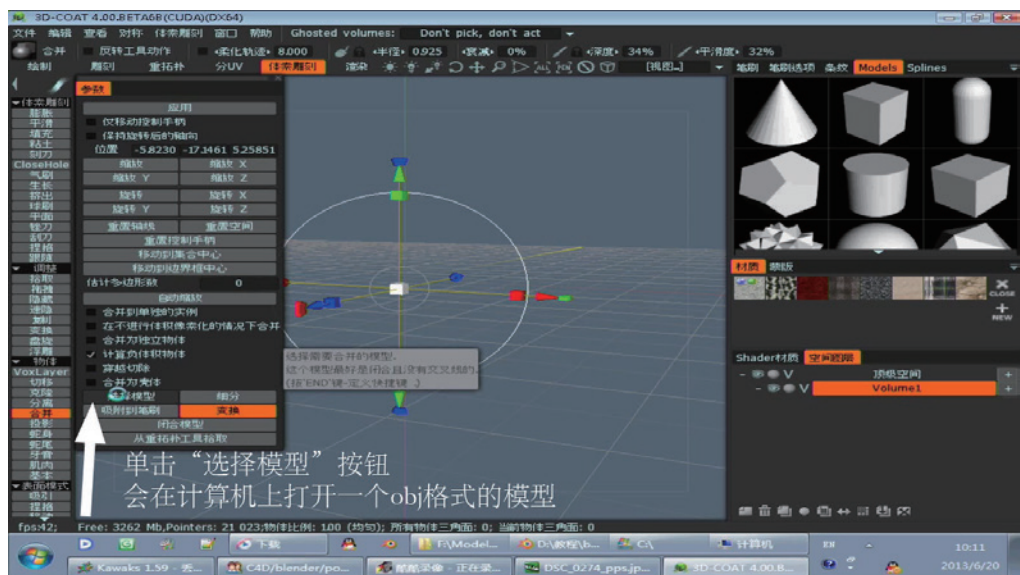


图 5-42 单击“选择模型”按钮，选择需要编辑的原始扫描 3D 模型

导入模型后，软件会显示该 3D 模型的三角形面片数，我们可更改数值以便对模型进行简化，如图 5-43 所示。

然后，我们就可以根据用户的照片，开始修正原始模型了，如图 5-44 所示。

如图 5-45 所示，软件中不同的笔刷有不同的作用，各自的效果需要分别进行实际体验。



其中，“牙膏”笔刷的作用是修补空隙，如图 5-46 所示。



图 5-46 “牙膏”笔刷的作用是可以修补空隙

当头发大体修复后，就可以接着添加细节了，如图 5-47 所示。

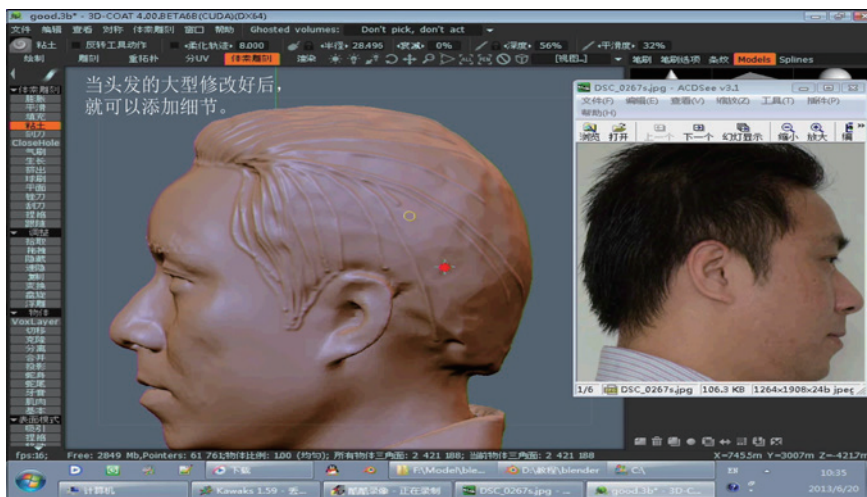


图 5-47 头发大体修复后，可以接着添加细节

当模型全部修正后，就可选择“文件”菜单的“输出”命令，并选择输出“物体”，然后保存为 3D 打印机可读取的 STL 格式即可，如图 5-48 所示。

头发修复后的多视角效果如图 5-49 所示，是不是跟参考照片的形状几乎一样了？

限于篇幅，本节无法给出详细的操作步骤。但在本书的网络下载栏目中，有一个时长 50 多分钟的完整视频教程，由 3D 建模专业人员录制，读者可仔细观摩。

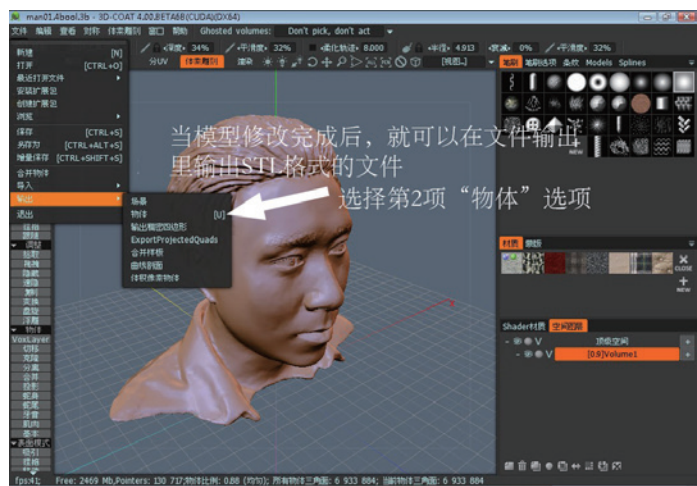


图 5-48 选择“文件”菜单的“输出”命令，保存为 STL 格式



图 5-49 头发修复后的多视角效果

另外两款雕塑软件 ZBrush/Autodesk MudBox 的功能与 3D-Coat 类似，在此不再赘述。ZBrush/Autodesk MudBox 功能更为强大，如图 5-50 所示，给出了一名创客使用 ZBrush 进行 3D 人体建模、光照渲染的例子。

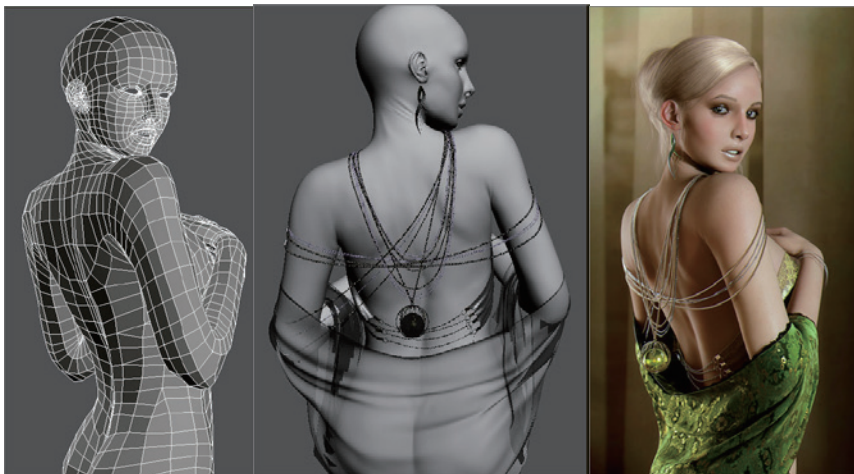


图 5-50 使用 ZBrush 进行人体建模

5.5.2 基于视觉计算自动修补发型

在视觉计算研究领域，头发建模是最困难的研究方向之一，目前还没有得到很好的解决。困难的根源在于：每个人的头发由近 10 万根头发丝组成（其中，黄种人有 10 万根；金色头发的白种人发丝较细，有 12 万根；红色的头发略粗，有 8 ~ 9 万根）。平均每平方厘米内就有约 150 根头发，这样直接导致了采集图片中每根头发丝的分辨率很低。此外，头发的材质表面不满足朗伯（Lambertian）反射模型，导致大多数主动光源 3D 扫描仪无法正常获取。更困难的是，头发丝之间彼此严重遮挡。

目前，对于头发的 3D 重建，采用视觉计算方法能取得不错的结果。用相机绕着客户旋转一圈，拍下若干张不同视角的图片，然后使用基于图像的建模（IBM, Image-Based Modeling）来生成发型。具体来说，通过提取图像中头发的朝向来构造光滑的矢量场，其中每根头发丝由一系列首尾连接的 3D 图形线段合成。然后根据多视角的内在几何约束，来重建出头发的 3D 形状。如图 5-51 所示分别是香港理工大学和浙江大学的研究小组所展示的发型重建和编辑效果。





图 5-51 基于图像的建模对发型进行 3D 重建和编辑(图片来源：香港理工大学、浙江大学)

有些“精益求精”的读者可能会说：上面的结果只是看起来大体相像，每根头发丝是用 3D 线段近似拟合的，并不是严格真实的，而我，希望把我脑袋上的 10 万根头发（至少是那些没被遮住的）一根一根地真实重建出来！好吧，国外还真有人这么做了。具体来说，将高精度单反相机放置在一个可前后移动的导轨上，录制一段各个景深上的视频图像。然后利用特征提取和视觉算法，我们甚至可以将每一根头发丝 3D 重建出来。如图 5-52 所示，其中下图是重建前后的头发丝对比，可以看出重建后的 3D 头发确实与 2D 视频中的真实头发相吻合。



图 5-52 利用特征提取和视觉算法 3D 重建每一根头发丝(图片来源：Cornell University)

上面的方法在一些较为理想的拍摄环境下（如高清摄像机、充分的光照条件、客户的头发干净整洁）可取得很逼真的效果，然而在实际的应用环境中却往往是随意的、不够理想的。计算也较为费时，无法满足照相馆“立等可取”的需求。实际上，对于目前 3D 打印机有限的打印精度而言，特别精细的头发重建也是不必要的，我们往往只需要获取到发型的轮廓和发束的走向即可，如图 5-53 所示。在这里，发型修补算法成功的关键在于要适应非充分光照条件、中低

精度的照片采集质量，以及尽可能减少用户的交互，以实现一个全自动、健壮的总体发型重建。其中要涉及图像中发型轮廓的特征提取、多视角几何约束下的三维重建、基于侧影轮廓线构造可见外壳（Visual Hulls）等技术，感兴趣的读者可访问“视觉计算研究论坛”（<http://www.sigvc.org/bbs>）了解更多内容。



图 5-53 适合 3D 打印的头发建模，只需发型轮廓和发束走向即可（为保护隐私，眼部已遮挡）

为了进一步提取细节，迪斯尼的研究人员提出了一种新的头发多视角风格化（这里可理解为“抽象化”）算法^[72]，将 2D 图像中的特征保持（Feature-Preserving）颜色滤波算法扩展到不规则的 3D 参数化域中，并引入与 2D 颜色风格化相一致的几何细节。图 5-54 展示了为不同的人分别采集多张不同视角的 2D 图像所重建出的 3D 头发细节。



图 5-54 迪斯尼研究人员提出的一种新的头发多视角风格化算法

还有些“差不多就行”的读者可能会说：其实不用那么较真，不用生成跟真人一模一样的发型，我还觉得自己的发型不好看呢，能换个别的发型不？当然可以！这其实也是最简单、最常用的解决方案。我们需要事先准备一个 3D 发型库，如图 5-55 所示，里面有做好的各式各样的 3D 发型。这时换发型就很简单：直接把之前的头发去掉，套上新发型后编辑合成一下，最后再重建成一个紧密的流形曲面以便输出到 3D 打印机。



图 5-55 基于 3D 发型库的编辑合成（图片来源：南加州大学）

5.5.3 Geomagic Studio：更通用的任意形状修补

下面我们就详细介绍一下 3D 照相馆必备的利器：Geomagic Studio，可对任意形状进行修补和平滑，其地位与现在 2D 照相馆常用的 Photoshop（PS）相当。

很多新手一开始不知道如何操纵一个 3D 模型，Geomagic Studio 定义的快捷键如下。

- **旋转 3D 模型**：按住并拖动鼠标中键（或 Ctrl + 鼠标右键）。
- **平移 3D 模型**：先按住 Alt 键，按住并拖动鼠标中键（或 Alt + 鼠标右键）。
- **缩放 3D 模型**：前后拨动鼠标中键滚轮（或 Shift + 鼠标右键）。

我们将一个物体用 3D 扫描仪扫描后，得到的原始数据是各个扫描侧面的点云数据。因此，首先需要将各个侧面拼接成一个整体，即将各个视图得到的点云合并到一个公共坐标系下，从而得到一个完整的模型。对于 3D 扫描流程不太清楚的读者，可回顾 5.2 节“3D 照相馆的核心技术：3D 智能数字化”。

比如我们分两次扫描了一个 3D 头像，分别得到了左半脸和右半脸的点云数据。将 3D 数据

通过左上角的“文件”菜单“导入”到 Geomagic Studio 后,在主界面左侧的“模型管理器”中,按住 Ctrl 键,用鼠标左键分别将两片点云选中,如图 5-56 所示。

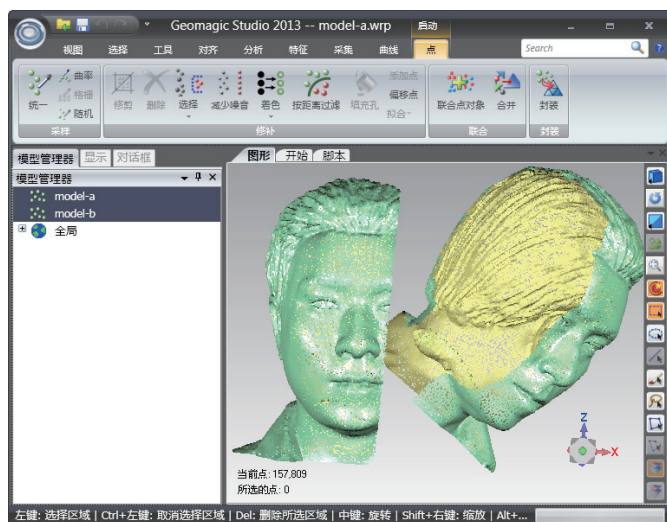


图 5-56 在“模型管理器”中将两片点云选中

然后单击“对齐”→“手动注册”按钮。一般选择“n 点注册”,配准的效果比“1 点注册”效果要好。在左侧窗格中指定要对齐的两片点云(一个选作“固定”,另一个选作“浮动”),如图 5-57 所示。

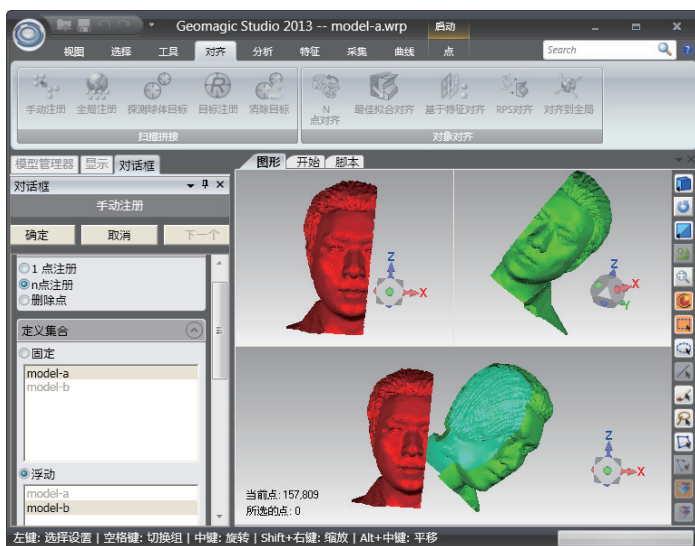


图 5-57 选择“手动注册”的“n 点注册”

“n 点注册”中的 n 点,至少要有 3 个点。因此,我们在两片点云数据的重叠处选择 3 对对应点,比如在左半脸中我们选择了鼻尖作为第 2 个对应点,那么在右半脸中也要选择鼻尖作为第 2 个对应点,如图 5-58 所示。

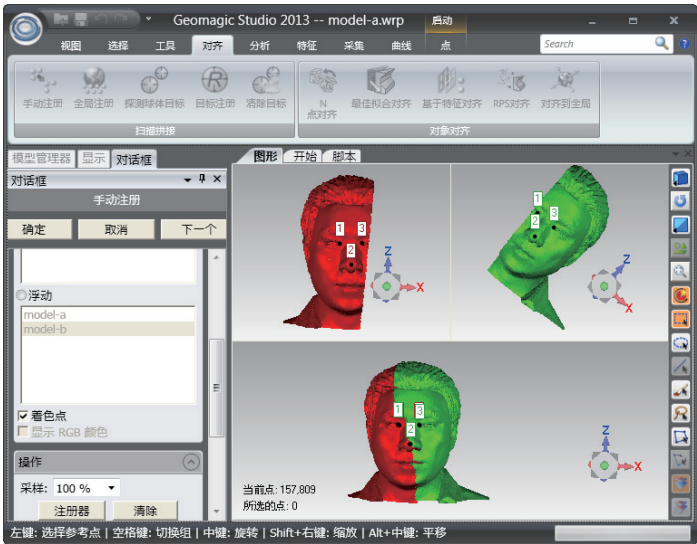


图 5-58 在两片点云数据的重叠处选择 3 对对应点



注意:如果两片点云大部分重合,或只能找到 1 个较为精确的共同特征点,这时可选择“1 点注册”。但一定要记得在图 5-57 右侧的两个预览小窗口中,把两片点云分别旋转到同一个朝向角度,然后再分别指定共同的特征点,比如鼻尖。

指定好特征点对之后,单击“注册器”按钮进行配准对齐。如图 5-59 所示,我们在界面左侧的下方可以看到误差统计,距离和偏差越小则代表精度越高,一般以满足实际需求为准。

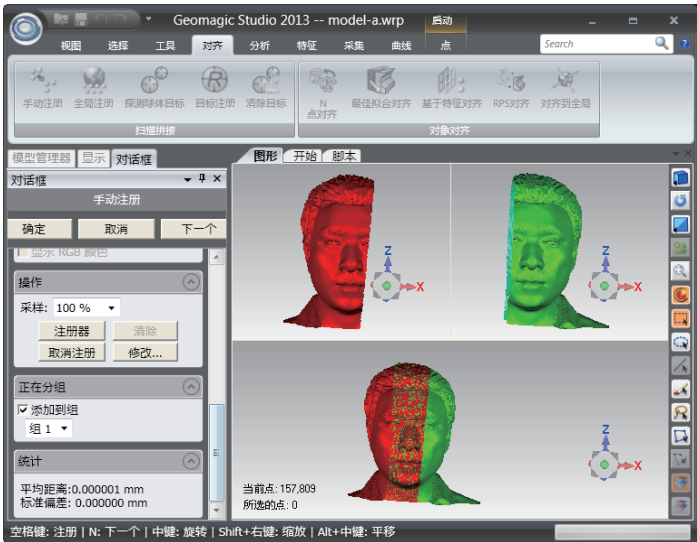


图 5-59 将两片点云配准对齐

手动注册之后,点云片之间的定位还会存在一定的误差,这就需要对点云进行全面的、整体的位置调整,即全局注册,以消除定位误差。单击“对齐”→“全局注册”按钮即可。

3D 扫描的原始数据经常会含有一些比较大的噪声点,可分别选择“点”→“选择”按钮下的“非

连接项”和“体外孤点”选项,然后按键盘上的 Del (Delete) 键将这些孤立的、没有连接的点删除。如果你还不放心,可进一步单击“点”→“减少噪声”按钮。

由于注册对齐后的点云分布并不均匀(比如重叠处的点就非常密集),这时可单击“点”→“统一”按钮进行统一采样。如图 5-60 所示,通过中间下方的提示可以看到,之前的 157 809 个点被均匀采样成了 116 174 个点。

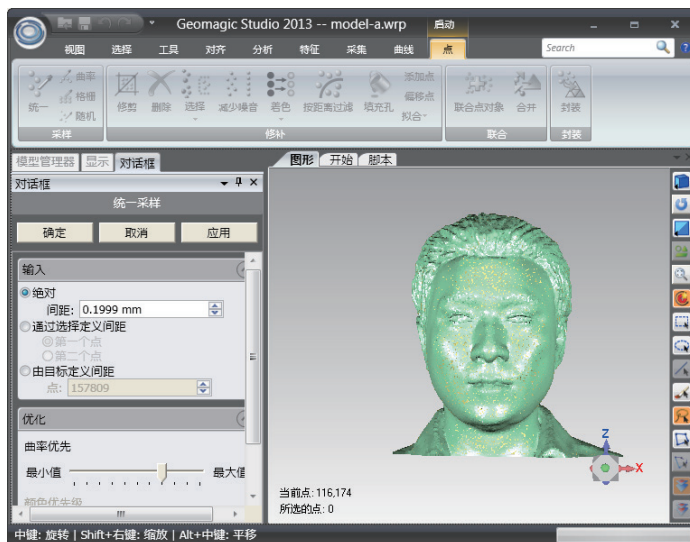


图 5-60 对注册对齐后的点云进行统一采样

OK, 下面我们就将对齐后的两片点云合并成一个统一的三角形网格。单击“点”→“合并”按钮,并在左侧的“高级”栏中选中“删除重叠”复选框,将右边文本框的数值放大 10 倍,比如原本是 0.292mm,则更改为 2.92mm,如图 5-61 所示。

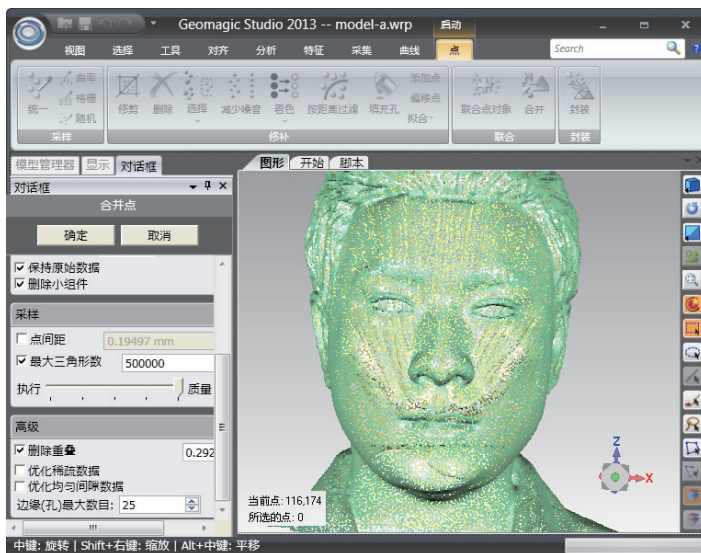


图 5-61 “合并”时“删除重叠”部分

单击“确定”按钮完成合并，点云就被转化为一个三角形面片模型了。如图 5-62 所示，通过中间下方的提示可以看到，之前的 116 174 个点现在变成了 239 188 个三角形面片。

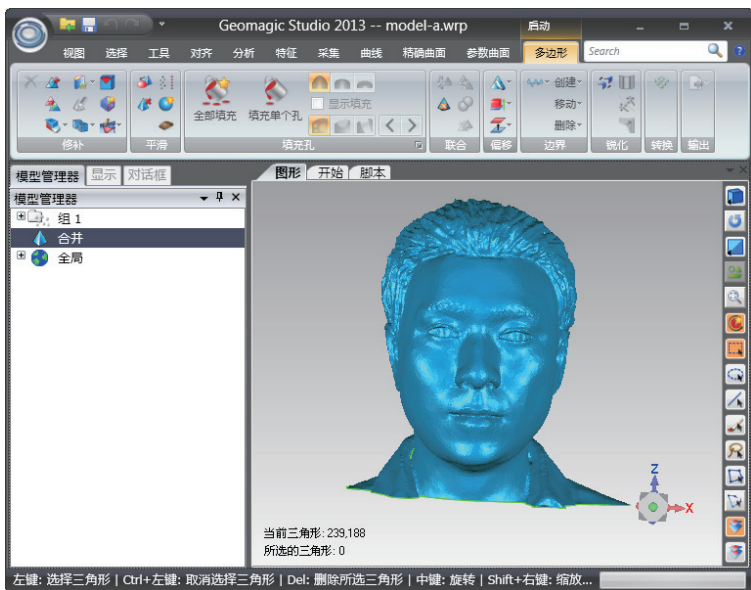


图 5-62 合并后将点云转化为一个三角形面片模型

有些心急的读者会说，现在就可以将三角形模型“另存为”STL 格式文件进行 3D 打印了吧？别急，这个三角形模型往往并不完美，有些地方会出现孔洞，如图 5-63 左边所示的土黄色部位。这时需要对其进行填补，操作很简单：单击“多边形”→“填充单个孔”按钮，然后把鼠标放在孔洞的边界，按下鼠标左键即可完成填补，如图 5-63 右边所示。然后单击“>”按钮定位到下一个孔洞进行填补。

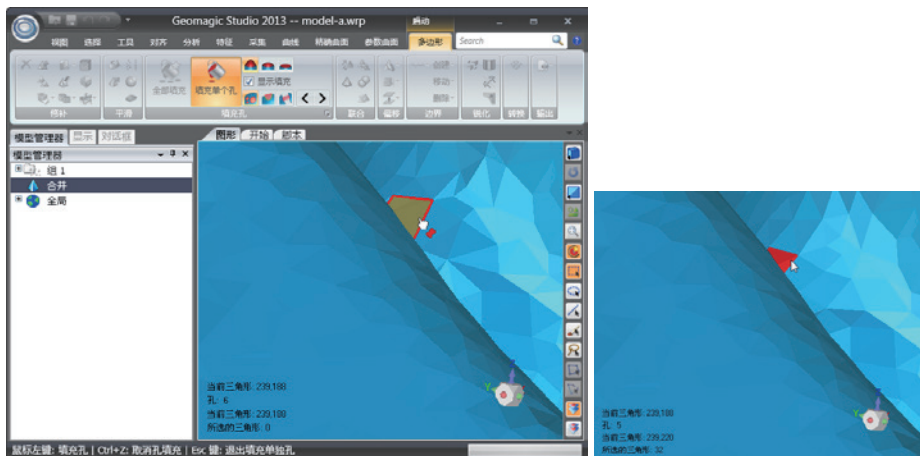


图 5-63 填充孔洞

一个一个找孔洞确实比较烦琐，你可以直接单击“全部填充”按钮自动将所有孔洞填补好。如果你还不放心，可进一步单击“多边形”→“网格医生”按钮，对整个模型网格做一次彻底

大体检和自动修正，如图 5-64 所示，网格医生会将有问题地方用红色标记出来。

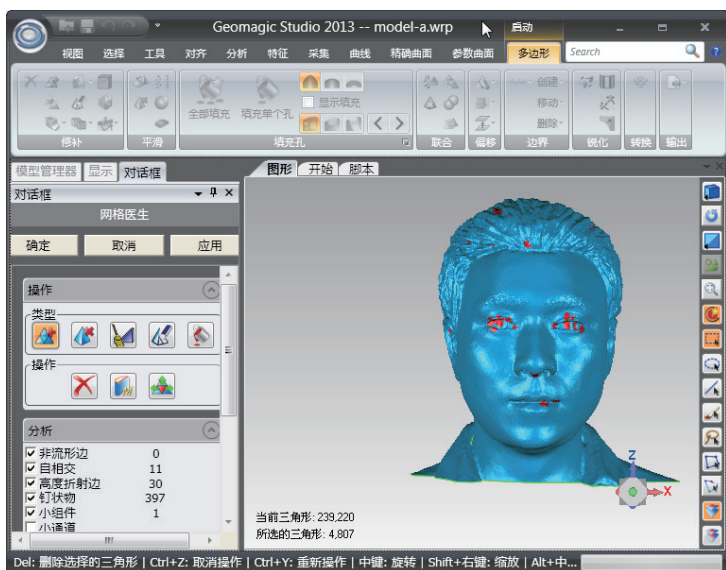


图 5-64 网格医生对整个模型网格做检查和自动修正

经过这么一折腾，3D 模型现在看起来非常不错，三角形面片的数目也变成 239 220 个了。你嫌它有点多，想精简一下，这时可单击“多边形”→“简化”按钮，输入希望的“目标三角形计数”，比如 159 018 个，单击“确定”按钮即可成功“瘦身”，如图 5-65 所示。

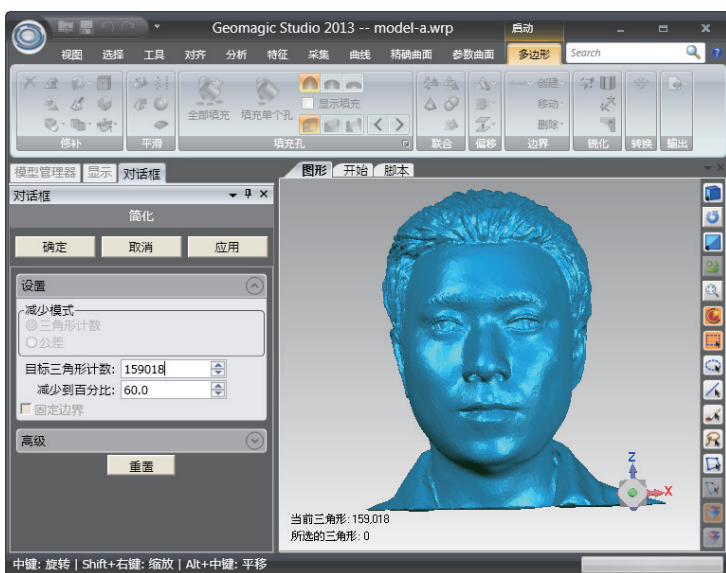


图 5-65 对三角形面片进行精简

有些顾客的脸上不是太光滑，比如有些小痘痘，3D 扫描后也将它们忠实地记录下来了。怎么办？这时可单击“多边形”→“砂纸”按钮，然后按住并拖动鼠标左键不停地打磨即可。如图 5-66 所示，打磨后是不是光滑了许多？

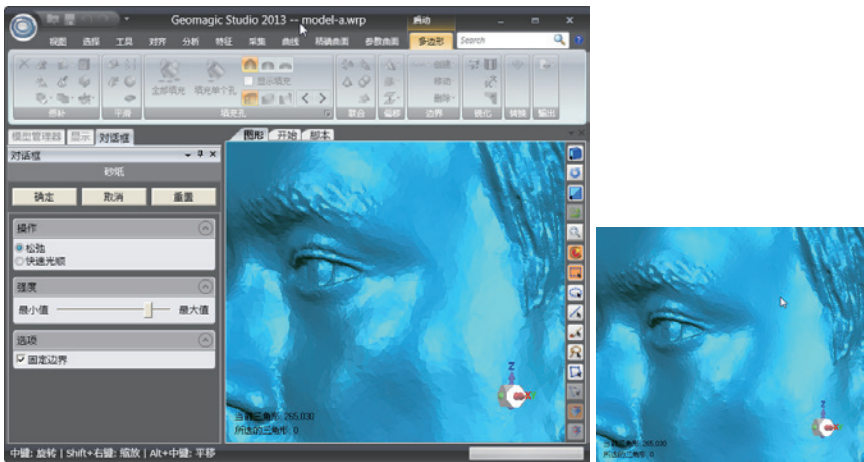


图 5-66 用砂纸进行打磨

如果有的顾客说，我脸上有块刀疤，砂纸总磨不掉，怎么办？Geomagic Studio 再一次展示了它温存贴心的一面：先把这块刀疤选中，然后单击“多边形”→“去除特征”按钮，OK，一切恢复如初，犹如婴儿般爽滑，如图 5-67 所示。

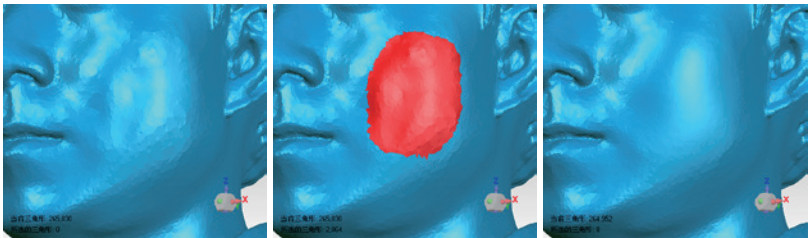


图 5-67 “去除特征” 获得局部平滑填充效果

有的懂行的女读者会说，最后能不能再打一次“粉底”？没问题！单击“多边形”→“松弛”按钮，即可对整张人脸进行全局的自动平滑，如图 5-68 所示。

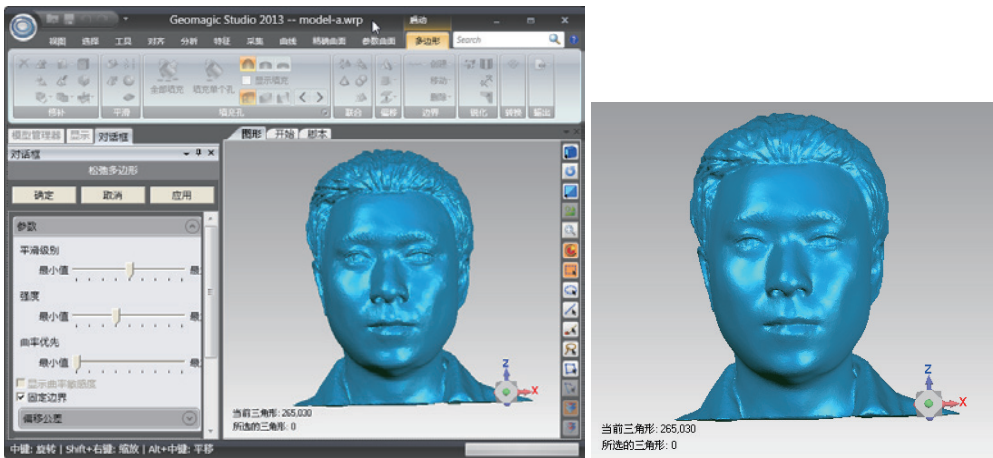


图 5-68 使用“松弛”功能对整个网格进行全局平滑

5.6 3D人脸表情形变与编辑

在3D照相馆中，顾客一般都是正襟危坐，表情严肃地保持一个姿势几分钟，直到3D扫描完成。在将模型3D打印之前，可能很多顾客又不甘心这么平淡，希望能对自己刚才的表情做一些修改和编辑。比如将刚才的双唇紧闭、面无表情，修改为明眸皓齿、回眸一笑。利用3D智能数字化技术，这是完全可以的！相关原理可参考第6章6.2.2节“个性特征的定位与匹配”，只不过这里将形状和纹理替换成了表情。具体来说，我们事先建立一个3D人脸表情数据库，里面含有各种各样的3D人脸表情。然后我们对这些表情进行智能化分析，提取出一些基（本）表情，如愤怒、恐惧、开心、惊讶、厌恶等。当然我们也可直接从数据库中手工指出含有这些基表情的3D形状（Blendshapes），以便直接告诉计算机这些语义信息。于是，人的所有表情都可表示为这些基表情的加权线性组合。比如，你可定制一个表情，其中含有10%的愤怒+20%的恐惧+30%开心+15%惊讶+25%厌恶。如图5-69所示，我们对3D表情进行了编辑，如改为愤怒、恐惧、开心等。

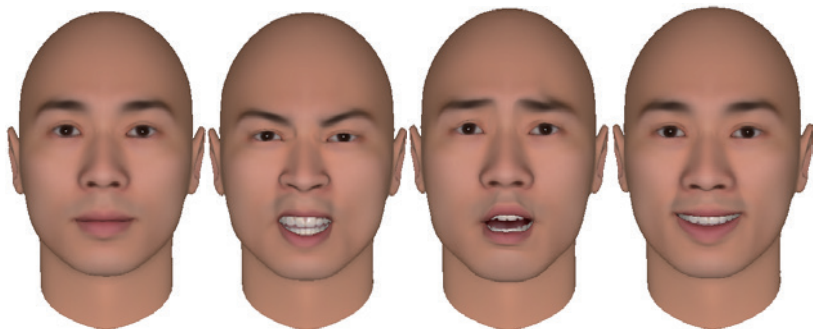


图 5-69 对 3D 表情进行编辑（愤怒、恐惧、开心）

然后，很多情况下人类的表情是很微妙的，如愤怒就分很多种（娇怒、狂怒、大怒、佯怒等），如果要用基表情混合出来，则需要反复地手工调节参数，比如“愤怒”的权值应该是80%还是85%，“开心”的权值应该是10%还是12%？更麻烦的是，一个复杂的表情往往由50~100个基表情组成，这就意味着你需要同时调节50~100个参数！人生苦短，因此，我们可换种方法。比如，事先采集一个海量的真实表情数据库^[41]，如图5-70所示，表演者把几乎可以做出的任何表情都做了一遍。注意贴在表演者脸上的标记点（Markers），用于动态捕捉人脸的3D表情。

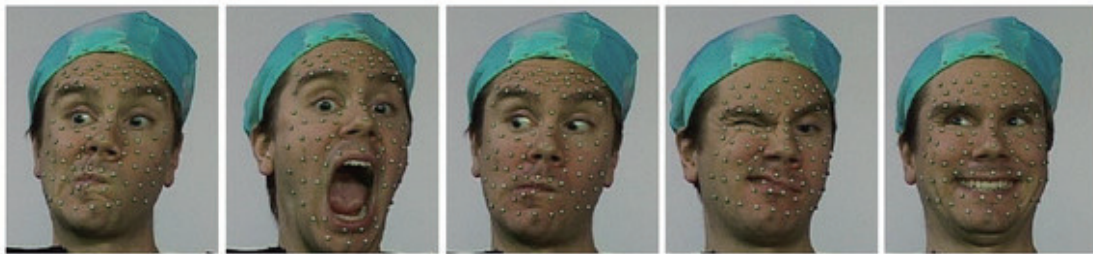


图 5-70 人脸表情的运动捕捉（图片来源：微软）

这时，用户就不再需要费劲地手工调参数了，只需轻轻选定某个参考表情，然后让计算机自动把它**形变迁移**（Deformation Transfer，也被称为**重定向** Retarget）到自己的脸上即可。如图 5-71 所示给出了一个表情迁移序列^[70]，把上方的一张胖脸的表情序列，忠实地迁移到下方的一张瘦脸上，使其做出相同的表情动作。



图 5-71 人脸表情迁移（重定向）（图片来源：MIT）



提示：Deformation Transfer 的本质思想非常简单，可用如下公式表达：

$$\min_{v_1 \dots v_n} \sum_{j=1}^{|M|} \|S_{s_j} - T_{t_j}\|_F^2$$

其中 S, T 代表彼此相对应的源（Source）/ 目标（Target）三角形面片上的源 / 目标形变。也即：目标形变应尽可能地与源形变相同。这个问题最终可归结为一个稀疏线性方程组的求解：

$$\min_{v_1 \dots v_n} \|c - A\tilde{x}\|_2^2。$$

前面介绍的方法，要么需要调参数，要么需要像在商场选衣服一样仔细选择中意的表情，都不太直接。那么，我们能否直接在 3D 人脸上操纵表情呢？就像小时候妈妈用手捏起你的左脸颊，你的左脸就会自动往上翘，并痛苦地做出相应的表情。也即，你只需交互地操纵某一个局部地方（如嘴角、脸颊、眼角等），则整个脸部的表情会跟着改变，同时又能满足你的操纵约束（比如嘴角上提到某个位置）。如图 5-72 所示，用户只需操纵 3D 人脸的嘴角和眼角部位（见中间图片的嘴角和眼角上的黑色操纵点），则整张人脸就会跟着变化表情。

由于操纵点个数非常少，而整个人脸表情可变化空间却又非常大，因此往往需要将表情空间降维约束到一个子空间中，以减少歧义。在第 6 章 6.2.2 节中我们将介绍：降维常采用 PCA 方法，以便降到一个线性子空间（Subspace）中。当然，更复杂的还可以考虑由多个局部线性子空间构成的非线性空间，以便得到一张更符合人类自然表情的人脸。在图中我们还可注意到，人脸的各个部位用不同的颜色分割出来，这样的好处是为了更加灵活地操纵，可以生成表情数据库中原本并没有但仍自然的表情（比如左脸很自然地笑，而右脸很自然地哭）。

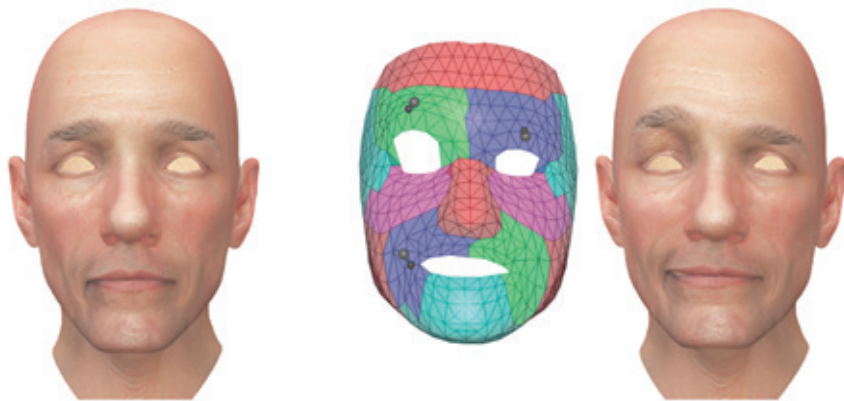


图 5-72 直接交互地操纵 3D 人脸表情(图片来源 : CMU)

有的用户可能还是会说，交互地直接操纵表情虽然直观，但如果我懒到连动手都不愿意，怎么办？没关系！你真的可以不用手，而直接用自己的脸做出相应的表情来驱动 3D 表情！这里需要用到视频人脸表情的实时检测和跟踪技术，如我们可以利用第 6 章 6.2.2 节的 AAM（Active Appearance Model，主动外观模型）技术对视频中的人脸进行实时定位，同时驱动 3D 模型做出和你一样的表情，如图 5-73 所示。

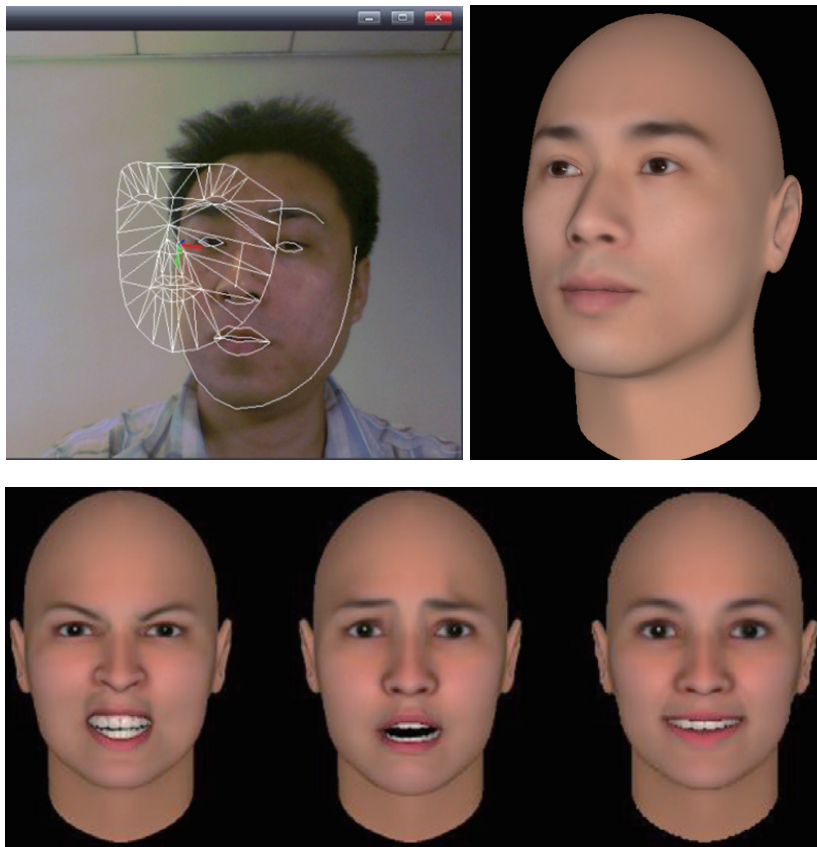


图 5-73 利用视频驱动 3D 人脸表情（上图为笔者本人）

我们对上面的人脸表情编辑方法做一个小结。我们倾向于采用直观的人机接口用于人脸表情的交互编辑，比如简单地拖动控制点和控制线条，或者直接用自己的人脸视频去控制。人脸表情编辑系统成功与否的关键在于要避免生成不自然的人脸表情。为了实现这个目标，可预先录制和采集一个真实人脸表情数据库。由于 3D 数据库的维数很高，处理起来非常复杂费时，一般都需要降维到子空间中去构造一个先验统计模型。这个先验模型的主要目的是限定生成的人脸表情应位于自然表情空间范围之内，这样就有效地防止了表情的失真。



扩展：我们可将三维人脸数据库整合成一个三维数据**张量**（Tensor） \mathbf{T} 。张量是**多重线性代数**（Multilinear Algebra）中的概念，是线性代数在更高维度上的泛化，也即张量是**向量**（一维）和**矩阵**（二维）在高维（三维及以上）的推广。利用张量 \mathbf{T} ，我们可以将三维人脸数据库中不同的三维顶点位置、不同的测试者、不同的表情变化这 3 个属性结合到一起，如图 5-74 所示。

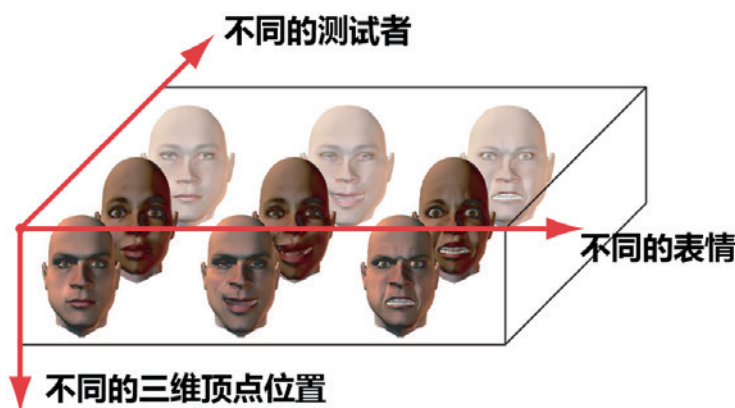


图 5-74 利用数据张量将三维人脸数据库中不同的三维顶点位置、不同的测试者、不同的表情这 3 个属性结合到一起变化（图片来源：MIT）

张量运算中一个特别有用的线性变换就是 N 维的**奇异值分解**（N-mode SVD），也即**张量分解**（Tensor Decomposition）。比如，我们将**数据张量** \mathbf{T} 沿着不同的测试者和表情这两个属性进行奇异值分解：

$$\mathbf{T} \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T = \mathbf{M}$$

$$\Rightarrow \mathbf{T} \cong \mathbf{M}_{\text{reduced}} \times_2 \tilde{\mathbf{U}}_2 \times_3 \tilde{\mathbf{U}}_3$$

其中**核心张量** $\mathbf{M}_{\text{reduced}}$ 称为**双线性人脸模型**^[73]，类似于矩阵奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ 中的 $\mathbf{\Sigma}$ ； $\tilde{\mathbf{U}}$ 通过去掉矩阵 \mathbf{U} 中冗余的后几列来得到，以获得高质量的近似。利用双线性人脸模型 $\mathbf{M}_{\text{reduced}}$ ，我们就可以生成数据库中任何一张带表情的人脸：

$$\mathbf{f} = \mathbf{M}_{\text{reduced}} \times_2 \mathbf{w}_2^T \times_3 \mathbf{w}_3^T$$

其中 \mathbf{f} 是得到的新的人脸； \mathbf{w} 是一个列向量，包含的元素是 $\tilde{\mathbf{U}}$ 中不同行线性组合的系数。

好，以上这么多种方法，终于可以满足 3D 照相馆顾客的人脸表情更换需求了。然而，人的进取心总是不断向上的。有的顾客会说，虽然我从小就有一条腿不太好使，但我一直梦想腾飞的感觉，我的 3D 打印人像能够替我做出跳街舞、翻筋斗云的动作来吗？答案是肯定的！跟人脸表情编辑的原理一样，我们也预先采集一个人体**运动捕捉（Mocap，Motion Capture）**数据库，里面有动作演员做出的各种高难度动作，接着也是在子空间构造一个先验统计模型，然后就可以把一个个很酷的 3D 动作（如图 5-75 所示）迁移到你的 3D 人像上了。



图 5-75 3D 运动捕捉数据的采集和迁移（图片来源：TU Munich）

5.7 直接全彩打印，还是单色打印再上色

但凡开 3D 照相馆的读者，都会反复思考应选择哪种方案：是直接全彩打印，还是单色打印再上色？因为这直接影响到开 3D 照相馆的成本，而且几乎是天壤之别。

下面我们就以北京和西安的两家照相馆为例，跟读者详细算一笔账。北京的一家 3D 照相馆采用的 Zprinter 650（现已合并到 3D Systems 公司的 ProJet 系列）是工业级别的彩色打印机。西安的一家 3D 照相馆采用的是桌面级打印机 MakerBot Replicator 2，打印出单色模型。两者的详细对比如表 5-1 所示。

表 5-1 全彩打印与单色打印对比

	北京 3D 照相馆	西安 3D 照相馆
设备类型	Zprinter 650	MakerBot Replicator 2
设备精度	0.1mm	0.1mm
设备价格	90 万元	3 万元
设备色彩	390 000 色	单色
设备级别	工业级别	桌面级别
使用材料	高性能复合材料、石膏粉	PLA
材料价格	高性能复合材料每千克成本约 3 000 ~ 4 000 元，石膏材料每千克成本约 500 元。此外还有黏结剂、墨水、胶水的费用	单色 PLA 或 ABS 打印材料每千克成本约 200 元
打印速度	3 ~ 5 小时 (参考模型高：110mm)	4 ~ 7 小时 (参考模型高：110mm)

无图无真相，下面我们看看实际的效果对比。如图 5-76 所示，左边是全彩打印的效果，右边是单色打印再上色的效果。



图 5-76 打印效果对比。左：石膏全彩打印的效果；右：单色打印再上色的效果（右图来源：立想 3D、成都 XYZ-3D）

可以看出，直接全彩方案的设备费用是单色方案的 20 ~ 30 倍，耗材成本也是 5 ~ 15 倍左右。具体地：

- 在产品细节上，如果顾客穿的是五颜六色且印有细小字母文字的服装，则直接全彩方案效果更好，因为后期手工上色无法描绘出细小的字母文字以及各种相间渐变的颜色。
- 如果顾客穿着的是具有大片相同颜色的服装，比如上身全黑、下身全蓝，且模型的尺寸比较大（如 20cm 以上），则单色打印再上色效果更好，因为可以用颜料调配出明亮的颜色；而直接全彩打印的色彩效果偏暗淡，颜色也有一定的失真。



思考：实际上，3D 人像模型只要求最外面的那一层是彩色的即可，而全彩工业级打印机则是里里外外都彩色打印，似乎有些浪费。笔者设想，以后也许可以先把身体表面各部分（如衣服、裤子）的彩色纹理按区域分成很多小块，分别打印在多张 2D 的转印纸上，然后分别往单色 3D 模型的各个区域粘贴，彩色图案就转印上去了。这个原理类似于计算机图形学中的纹理贴图（Texture Mapping）。

那么，到底应该如何选择呢？根据前面 5.1.3 节“3D 照相馆赢利模式的探讨”，普通用户可先选择单色打印再上色的方案，因为这样设备总成本可控制在 5 万元以内。不建议普通个人购买几十万的 3D 打印设备，不仅维护成本高，而且还要承受设备隔几年就要更新换代导致贬值的风险。此外，通过“智能云网”的服务模式，完全可以将扫描好的 3D 模型委托给专业的 3D 打印厂商进行全彩打印。这样，两种方案实际上可同时运作，可谓鱼和熊掌兼得。

5.8 3D打印数字化设计技巧

在第4章4.2节“3D智能数字化设计技术”中，我们已经介绍了3D数字化设计的原理。相信你已经对实体建模、曲面建模等概念有了清楚的了解。下面，我就带大家亲手操作一遍，通过一些实例介绍一下3D数字化设计的技巧。

5.8.1 3DS Max 建模用于3D打印

以下是一个关于实体建模（Solid Modeling）的基础教程，使用Autodesk 3DS Max进行三维设计并输出适用于三维打印的STL文件。

在本教程中，我们要设计一个简单的钥匙链挂件，涉及实体创建、二维转三维、布尔运算、组合对象、掏空对象等操作。如果你之前从未使用过任何3D建模软件，本教程会通过几个步骤告诉你做些什么。

1. 打开3DS Max，选择“Create”→“Shapes”→“Text”命令，如图5-77所示。

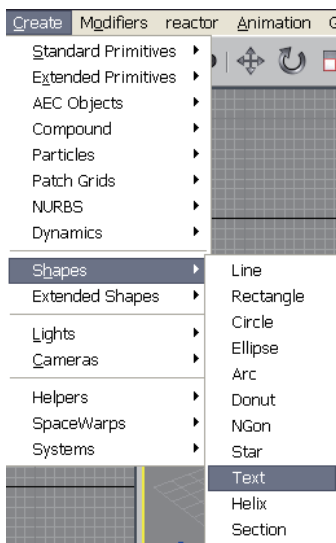


图 5-77 选择“Create”→“Shapes”→“Text”命令

2. 在屏幕某处按下鼠标，生成默认的文字“MAX Text”。然后在主界面右侧，在“Parameters”一栏中修改文字的各种参数。如图5-78所示，在数字输入框中设置字体的“Size”为14.0，然后在“Text”文本框中将默认文字“MAX Text”修改为“shapeways”，这样一个英文单词shapeways就出现在窗口中了。

3. 我们使用“Extrude”（伸出）命令让这个二维的文字变成三维的。选择“Modifiers”→“Mesh Editing”→“Extrude”命令，然后也是在主界面右侧，对“Extrude”命令进行参数设置，如图5-79所示。将“Amount”设为3.0、“Segments”也设置为3，把“Capping”栏中的“Cap Start”和“Cap End”复选框都选中，以确保这个三维文字是闭合的。

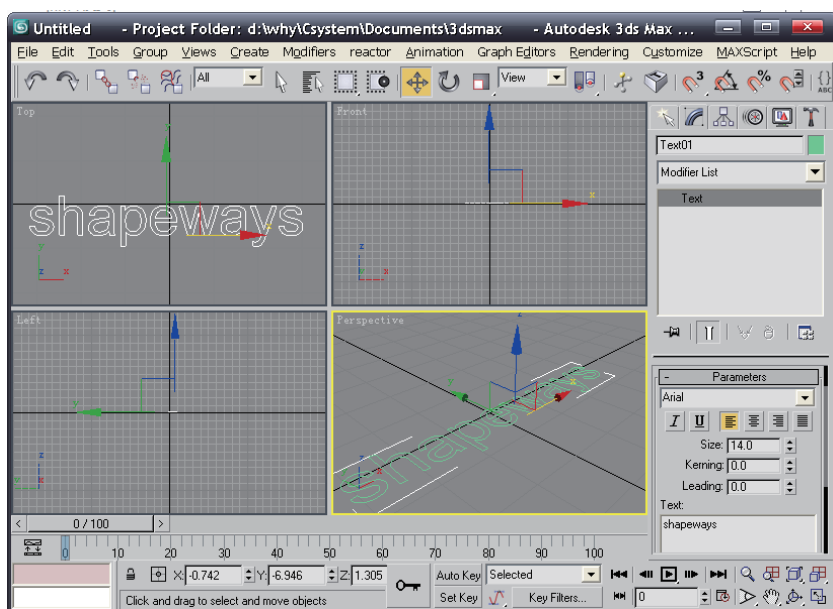


图 5-78 输入一个英文单词“shapeways”

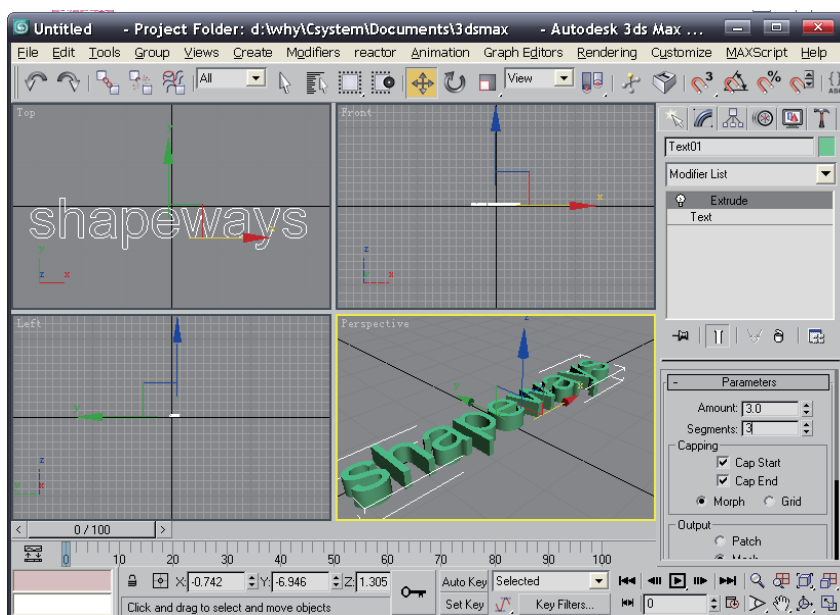


图 5-79 使用“Extrude”（伸出）命令让二维文字变三维

4. 接下来是创建一个长方体。选择“Create”→“Standard Primitive”→“Box”命令，按下鼠标在屏幕上拖拉生成一个长方体形状，然后在主界面右侧设置这个长方体的精确参数：“Length”为 12.0、“Width”为 66.0、“Height”为 3.0，如图 5-80 所示。如果这个长方体与文字 shapeways 不在同一平面上，可单击工具栏中的“移动”按钮进行调整。

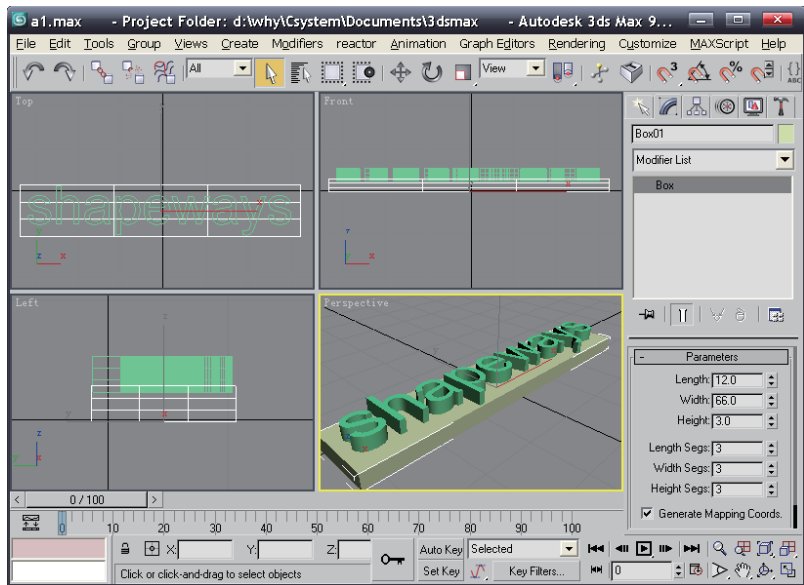


图 5-80 创建一个长方体

5. 使用布尔命令将这两个单独的物体合成一个物体。先选中长方体，再选择菜单中的“Create”→“Compound”→“Boolean”命令，在主界面右侧的布尔操作“Operation”栏中选择“Union”进行相加运算，然后单击“Pick Operand B”按钮并选择之前创建的文字。如图 5-81 所示，单词 shapeways 和长方体就合并在一起了。

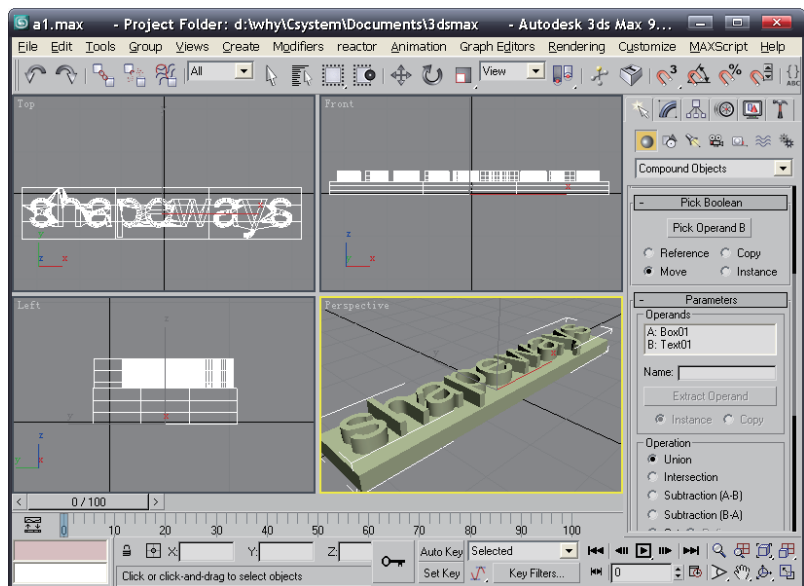


图 5-81 通过布尔加法运算将两个物体合并在一起

6. 接下来再举一个布尔减法的例子。我们要在这个印有单词 Shapeways 的钥匙链挂件上，穿一个小的圆柱孔洞，以便让吊绳穿过。首先选择“Create”→“Standard

Primitive” → “Cylinder” 命令创建一个半径为 1.8，高度为 5 的圆柱体。然后使用工具栏中的移动工具，将这个圆柱体移动到刚才合并后的模型上，如图 5-82 所示。

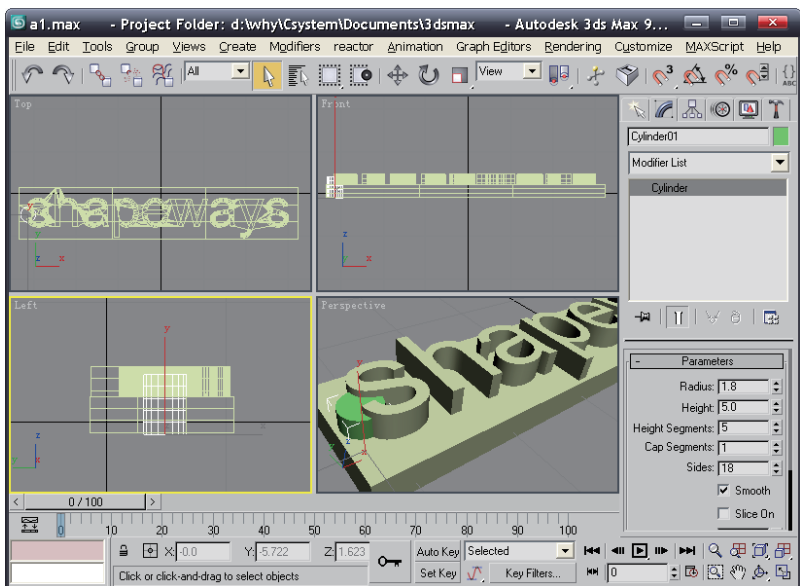


图 5-82 创建一个圆柱体并将其移动到已有模型上

下面我们将进行减法运算，而不是之前的加法运算。先选中刚才合并后的模型，然后同样也是选择 “Create” → “Compound” → “Boolean” 命令,但这次在主界面右侧的布尔操作 “Operation” 栏中选择 “Subtraction (A-B)” 相减运算，然后单击 “Pick Operand B” 按钮并选取圆柱体，这样就在合并后的模型中间挖出一个圆柱孔洞，如图 5-83 所示。

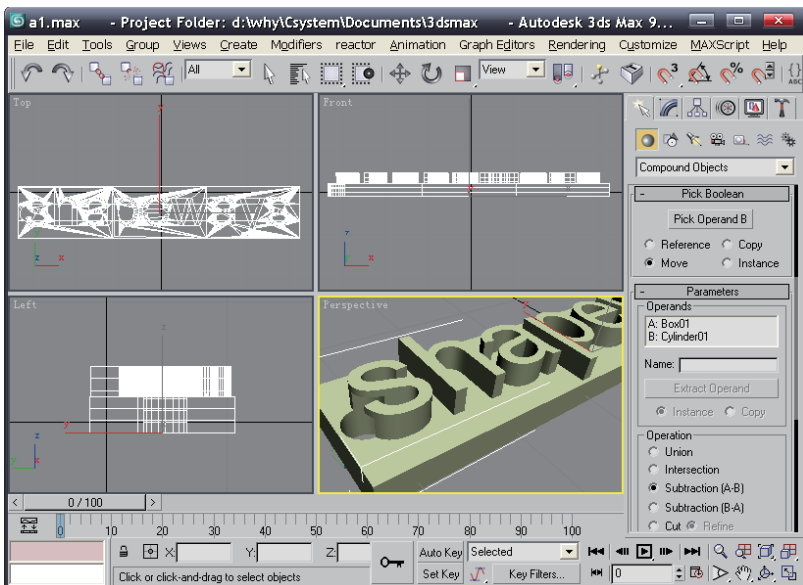


图 5-83 通过布尔减法运算将两个物体相减

7. 最后，选择“File”→“Export”命令将这个 3D 模型导出为 STL 格式的文件，就可以进行 3D 打印了。在导出数据的时候，请务必保证单位是毫米。因为 STL 文件有时是不包含单位信息的，所以需要指定单位是毫米，而不是米，这样在计算耗材价格的时候才不会弄错。

5.8.2 Netfabb/Magics：修正你的 STL 打印文件

在 3D 扫描或 3D 设计好一个 STL 文件之后，为了确保打印成功，我们需要在打印前先检查一下这个 STL 文件是否存在问题。下面我们就介绍使用 Netfabb 软件对 STL 文件进行各种操作。

如图 5-84 所示，在 Netfabb 中打开一个 STL 文件后，你可以在窗口的右下角查看这个 3D 模型的基本信息。

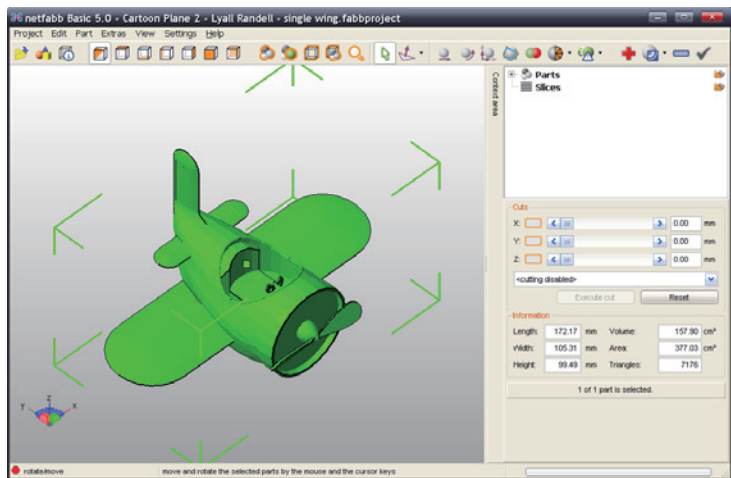


图 5-84 Netfabb 主界面

放大后的基本信息如图 5-85 所示，包括这个 3D 模型的长 / 宽 / 高、体积、总面积，以及三角形面片的总数量。

我们可单击工具栏中的相应按钮对 3D 模型进行平移、旋转和缩放，如图 5-86 所示。

Information					
Length:	172.17	mm	Volume:	157.90	cm³
Width:	105.31	mm	Area:	377.03	cm²
Height:	99.49	mm	Triangles:	7176	

图 5-85 模型的基本信息

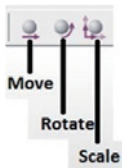


图 5-86 工具栏上的平移、旋转和缩放按钮

此外，工具栏中还有分析、修复、测量按钮，如图 5-87 所示。分析（Analyze）按钮可以提供关于 3D 模型的详细统计信息，如告诉你模型中是否有洞、边界边缘、翻转的三角面或者错误的边。修复（Heal）按钮是 Netfabb 的重要功能，可针对以上错误进行修复。测量（Measuring）按钮允许你测量两个面之间的距离，还可以测量最小壁厚。



图 5-87 工具栏上的分析、修复和测量按钮

我们先介绍如何对模型进行分析诊断。单击分析按钮，则 Netfabb 会将模型中有错的地方自动用红色标出来。但如果模型的绝大部分都被标为红色（99% 或更多，如图 5-88 所示），则可能是法线指向错误所导致的，这会让 3D 打印机无法判断出是模型的内部还是外部。解决办法是：先用鼠标选择模型，再选择“Part”→“Invert part”命令对法线进行翻转，以修正到正确的朝向上。

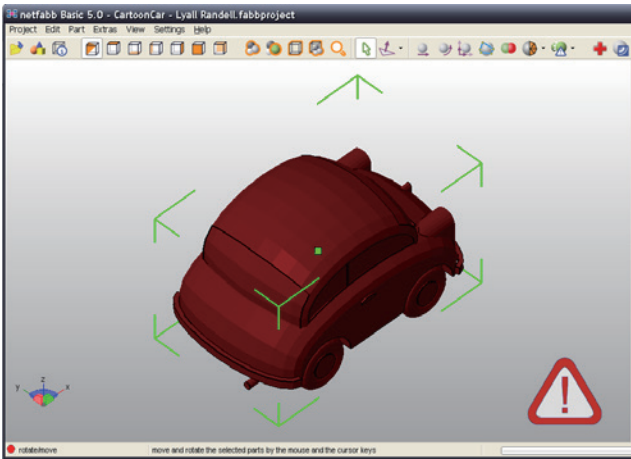


图 5-88 法线指向错误的 3D 模型

下面我们就开始介绍如何修复一个错误的 3D 模型（在错误状态下，模型无法被 3D 打印成功）。如图 5-89 所示，单击工具栏中的修复按钮（红十字），在主界面右侧可以看到设置栏。

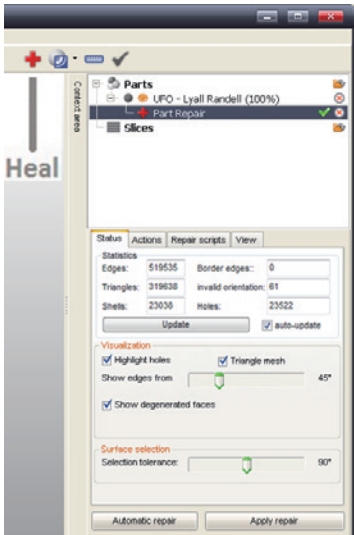


图 5-89 修复一个错误的 3D 模型

我们首先查看模型当前的统计信息，如图 5-90 所示。可以看出，当前 invalid orientation（无效方向）的数目为 61，Holes（孔洞）的数目为 23 522，真的是千疮百孔啊！让我们修复它吧！请注意在设置栏选中“auto-update”复选框，以便在修复后自动更新模型统计信息。

你只需轻轻单击设置栏左下角的“Automatic repair”按钮，就准备自动修复了。如图 5-91 所示，

你可以选择默认修复或者简单修复模式。推荐使用默认修复模式，这将会使用 Netfabb 提供的所有修复功能。

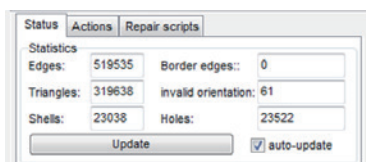


图 5-90 模型的详细统计信息

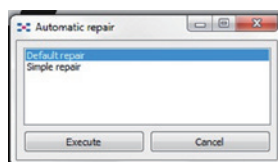


图 5-91 选择修复模式

修复完成之后，3D 模型的统计信息将会随之改变，如图 5-92 所示。可以看到，“Border edges”（边缘边）的数目已为 0，“invalid orientation”（无效方向）的数目也为 0，“Holes”（孔洞）的数目也为 0，“Shells”（壳壁）的数目为 1。完美！

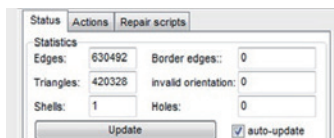


图 5-92 修复后的统计信息

记住，刚才的修复操作只是作为效果预览，还没有来真的。所以还需要再单击一下设置栏右下角的“Apply repair”按钮，将修复真正应用到原始错误模型。

好了，一切搞定！将修复好的 3D 模型导出到打印机吧。选择“Part” → “Export Part” → “as STL (binary)”命令，输入一个新的文件名保存即可。

除了 Netfabb 软件，Materialise 公司（著名的 i.materialise 在线打印服务便是该公司的业务）的 Magics 软件提供了一个更专业化的 STL 编辑处理平台，能够在几分钟内修复具有瑕疵的 STL 文件、创建打印支撑、生成切层，为接下来的 3D 打印做准备，如图 5-93 所示。操作原理类似，这里就不再赘述了。

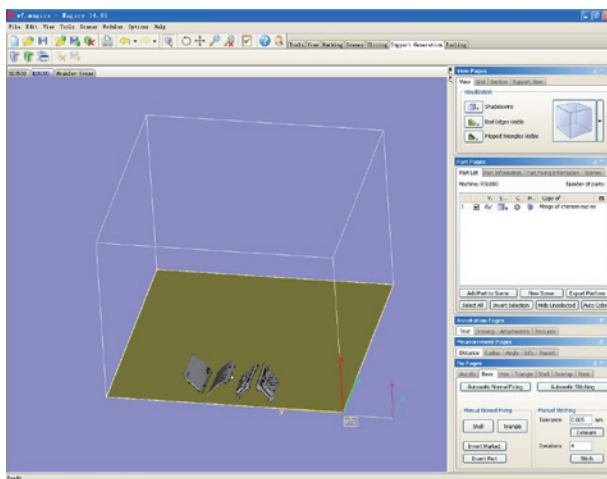


图 5-93 Magics 软件对 STL 进行编辑处理

5.8.3 使用 AccuTrans 3D 转换 3D 文件格式

在第 4 章中我们已经介绍了各种各样的 3D 数字化设计软件。然而，不是所有的 3D 软件都能够导出 STL、Collada 或 X3D 文件格式，或者它们导出的文件在进行 3D 打印时会出现错误。在这种情况下，可用文件格式转换工具 AccuTrans 3D 进行转换。此外，类似的工具还有 Deep Exploration、3DTransVidia、3D Object Converter 等。

启动 AccuTrans,选择“File”→“Open (all known formats)”命令,打开需要转换的 3D 文件。在转换格式之前，我们首先检查一下模型是否是水密的 (Water-tight)，也即不漏水的，3D 打印机要求流形网格数据中不能存在破洞。选择“Tools”→“Check for Water-tight Meshes”命令即可，如图 5-94 所示。

如图 5-95 所示,如果你在弹出的对话框中看到“OK”，那么就可安然无忧,进行下一步的操作。如果没有，则返回你的 3D 设计软件中检查网格数据中哪里存在破洞，并修补它。当然，你也可以使用 5.8.2 节介绍的 Netfabb 或 Magics 软件进行修补。

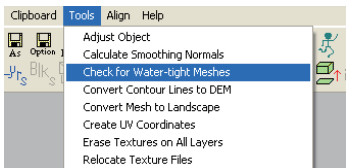


图 5-94 检查模型是否是水密的 (Water-tight)

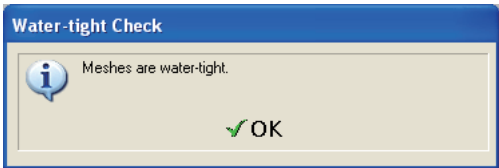


图 5-95 检查模型水密 (Water-tight) 成功

现在将模型文件转换为打印机可识别的格式，选择“File”→“Save with options...”。如图 5-96 所示，弹出了一个设置对话框。你对另存后的模型大小尺寸进行缩放（但请不要超出打印机的最大打印尺寸），只需在“Output Scale Factor”栏中输入合适的比例数值（见图中的①），则新的尺寸将会显示在“Scaled Object”栏（见图中的②）。

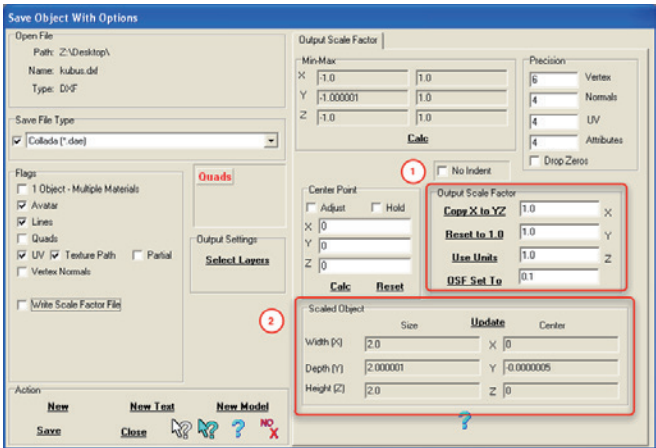


图 5-96 在转换文件格式前，对模型的大小尺寸进行缩放

单击“Save”按钮，可将你之前的模型文件另存为 Collada (*.dae)、STL、3DS、DXF、MA、OBJ 等几乎所有主流格式。当然，为了输出到 3D 打印机，你可以选择 STL 文件格式。

第6章

视觉计算：构建3D打印的杀手级应用

《论语·卫灵公》有云：“工欲善其事，必先利其器”。3D 打印要想真正走进千家万户，进入到人们工作生活的方方面面，必须要有杀手级的应用（Killer Application），给用户带来有价值的体验，让用户愿意为它不心疼地花钱。正如 iPhone 手机，虽价格昂贵，然而这并不妨碍 iPhone 成为大街上最流行的机型之一，根本原因就在于 iPhone 有强大的 App Store（应用程序商店），里面有许多杀手级的应用软件，让你每天都离不开它。前面章节我们已提到过，智能数字化与 3D 打印密不可分。而本章将要详细介绍的视觉计算，正是智能数字化的重要组成内容。掌握好它，让它成为 3D 打印的“利器”，构建一个个让用户爱不释手的杀手级应用出来。

6.1 视觉计算：计算机视觉与计算机图形学的融合

视觉计算本身是一门很新的学科，它是计算机视觉和计算机图形学发展到高级阶段的交叉融合。下面我们就分别对这几个学科做介绍。

计算机视觉（Computer Vision，简称 CV）是一门研究如何使计算机“看”的科学，更进一步地说，就是指用摄影机和计算机代替人眼对目标或环境进行感知、识别、跟踪和测量。人类的感官信息中，大多数是来自于视觉的。



提示：实验心理学家赤瑞特拉的著名心理实验指出：人类获取的信息 83% 来自视觉，11% 来自听觉，3.5% 来自嗅觉，1.5% 来自触觉，1% 来自味觉。

我们也可以把计算机视觉视为人工智能的一个分支。从这个角度来讲，可以认为计算机视觉的目的就是利用计算的手段来处理人类的视觉信息和实现对实际三维场景的智能理解。计算机视觉领域与图像处理、模式识别、射影几何、统计机器学习等学科密切相关。近年来，与计算机图形学等学科也有着很强的联系。



提示：模式识别用于从特征空间到类别空间的变换，通俗地说，就是自动将物体分类，如识别出这张照片拍的是鹿，那张照片拍的马。具体来说：根据从图像抽取的统计特性或结构信息，把图像分成给定的类别。研究内容包括特征提取（参见 6.2.1 节）、特征选择（参见 6.11.1 节）、分类器设计（参见 6.4）等。在计算机视觉中，模式识别技术常用于图像中某些部分（如分割区域）的识别和分类。

最早的且目前仍具有巨大影响的一种计算机视觉理论框架是由 MIT 教授 David Marr(大卫·马尔, 1945—1980 年) 在 20 世纪 70 年代末期提出的, 在他看来, 计算机视觉系统的输入是现实世界的二维图像, 而输出应该是基于 3D 表示的定性的和定量的场景理解。在 David Marr 英年早逝之后, 研究人员又相继提出了 Active Vision (主动视觉)、Purposive Vision (目的视觉)、Qualitative Vision (定性视觉) 等理论框架, 但这些新框架并没有代替 Marr 框架, 而是完善和丰富了 Marr 框架。国内从事计算机视觉研究的代表性机构有中国科学院自动化研究所的模式识别国家重点实验室, 自 1987 年成立以来对计算机视觉进行了系统的研究, 在计算理论框架、早期视觉处理、摄像机定标、三维结构重建、视频与医学图像理解等方向取得了一系列创新成果。

下面是计算机视觉的典型应用。

- 从一幅图像 (或一系列图像, 即视频) 中自动提取、分割感兴趣的物体 (例如, 提取人的面部)。
- 从多幅图像或序列中自动提取场景的三维信息, 如从几幅图片中实现对人体 / 人脸的三维重建。
- 在图像序列中自动跟踪有意义的移动物体 (如跟踪停车场中可疑的人的去向)。
- 从数字图像数据库中根据图像的视觉特征实现检索 (如从犯罪记录库中查找特定的嫌犯人脸图像、指纹图像、虹膜图像)。
- 根据摄像头抓取的实时信息进行交通监管。

下面我们介绍**计算机图形学, Computer Graphics, 简称 CG**。CG 这个简称相信更为人所熟知, 主要目的是利用计算机产生令人赏心悦目的三维真实感图形。例如, 我们看到的很多好莱坞大片 (阿凡达、变形金刚、钢铁侠等) 所展现的逼真炫酷效果就是用 CG 合成的。为此, 首先要对图形所描述的场景进行**几何建模 (Modeling)**, 再用某种光照模型, 计算在假想的光源、纹理、材质属性下场景的光照**渲染 (Rendering)**效果。计算机图形学的研究内容非常广泛, 如图形硬件、图形标准、图形交互技术、光栅图形生成算法、曲线 / 曲面造型、实体造型、真实感图形计算与显示算法, 以及科学计算可视化、计算机动画、自然景物仿真、虚拟现实等。国内从事计算机图形学研究的代表性机构有浙江大学的计算机辅助设计与图形学国家重点实验室。

当计算机视觉遇见计算机图形学又如何? 不管你相不相信爱情, 碰撞的火花诞生了。**视觉计算 (Visual Computing, 简称 VC)**主要研究利用计算机对视觉媒体数据 (包括 2D 图像、3D 模型、视频等) 进行获取、分析、合成、智能感知、可视化、交互和操纵, 其横跨计算机科学、数学、物理和认知科学。形象地说, **视觉计算既有计算机图形学的逼真炫酷效果, 同时又兼有计算机视觉的智能感知, 集“美貌与智慧”于一体**。国内从事视觉计算研究的代表性机构有中国科学院、浙江大学、清华大学、北京大学、北京航空航天大学等。此外还有一个专业的联盟组织“视觉计算特别兴趣研究组联盟” (SIGVC, Special Interest Group on Visual Computing), 网址为: <http://www.sigvc.org/>。该学术组织主要从事计算机视觉、计算机图形学和图像视频处理的前沿研究, 致力于国际一流的科研成果产出。感兴趣的读者还可经常访问“视觉计算研究论坛” (<http://www.sigvc.org/bbs>) 以了解更多内容。

计算机视觉、计算机图形学、视觉计算, 这三者之间既有区别, 又有联系。

- 计算机视觉是给定图像来推断场景特性, 实现的是从图像到场景的变换。即从二维图像

数据中分析提取场景的信息，包括三维结构、运动检测、物体识别等。

- 计算机图形学是给定关于场景结构、表面反射特性、光源配置及相机模型的信息，最后生成图像。从某种意义上说，计算机图形学是计算机视觉的逆问题。
- 而视觉计算是个更广义的学科，它包含了计算机视觉、计算机图形学、虚拟现实和可视化，也可以看作是这些领域在 3D、智能感知、人机交互上的交叉融合。

此外，表 6-1 对计算机图形学（CG）、计算机辅助设计（CAD）、3D 智能数字化（即视觉计算 VC）进行了详细的分析比较。

表 6-1 计算机图形学、计算机辅助设计、3D 智能数字化的分析比较

	计算机图形学 CG	计算机辅助设计 CAD	3D 智能数字化（视觉计算 VC）
数据表征	曲面	实体（CSG）	曲面与实体
制造形式	不制造	减材制造	增 / 减材制造
形状复杂度	不规则的曲面形状（一般要求流形曲面）	规则的体形状	不规则的曲面 / 体形状（支撑结构、中空、内嵌）
颜色纹理	表面纹理颜色	无纹理	表面纹理 / 体颜色
交互方式	2D 鼠标	2D 鼠标	2D 鼠标、笔画、体感、脑力、3D 鸟标
自动化程度	手动设计	手工设计 用于大规模批量生产	（半）智能化设计 用于大规模批量定制
物理力学	不考虑（除非物理模拟），因为应用场景为虚拟的比特世界	不考虑，因为加工时可用夹具等进行辅助	需要考虑重力、黏力
时间维	考虑（连续动画，至少需 25 帧 / 秒）	不考虑	可考虑（关键帧，离散动作）
应用时效性	短暂，用户一般只会观看一次动画和特效	长期功用，但大规模量产的工业产品千篇一律、一般不具反复观赏价值	个性化定制的实体打印作品，具有私人化特点，值得长久保存、反复观赏

6.2 3D打印“批量定制”的智能实现

我们在前面章节中反复强调过，3D 打印的优势就在于“批量定制”，以实现个性化制造。然而，这种追求高附加值的个性化定制，之前都是以较大的手工工作量为代价的，尤其是当需要“大批量定制”时。为提高定制效率，智能数字化技术将发挥关键的作用。比如，需要为一万名用户定制个性化的眼镜、服装、帽子、鞋子，如果使用人工逐一为每位用户进行手工测量和手工设计，工作量和成本都将变得不可接受。而应用智能数字化技术，如采用视觉计算方法，利用摄像头自动采集、分析提取每位用户的体貌个性特征，并自动根据视觉美感进行形状设计、颜色肤色搭配等，可极大地缩减定制周期。

下面（包括本章！），我们将开始我们的“技术控”之旅，希望你们能够 hold 住，在这期间我会不断地停下来等你，永不会放弃你，会反复用最通俗的话向你解说清楚，希望你能坚持到最后的胜利，必将有巨大的收获！

6.2.1 个性特征的描述与检测

所谓“特征 (Feature)”，是指“可以作为标志的显著特点”。而视觉计算中常说的图像特征，是指可对目标图像区域进行辨识的形状、纹理、颜色等特性，如角点 (Corner)、边缘就是典型的图像特征。再比如，你最喜欢她笑起来眼角微微翘起的样子，OK，眼角就是一个典型特征！当然，酒窝、美人痣、疤（如果你真欣赏她的话）什么的也都算。

图像特征描述是计算机视觉研究领域的一个基本问题，在寻找两幅相关图像中的对应点以及物体特征描述中有着重要的作用。它是许多方法的基础，因此也是目前视觉研究中的一个热点，每年在视觉领域的顶级国际会议 ICCV/CVPR/ECCV 上都有高质量的特征描述论文发表。

在介绍图像特征之前，我们首先仔细看一下什么是图像。我们平常数码相机、手机拍摄的照片就是图像！先来看图 6-1 中左边的彩色图像，可以看出一张图像是由很多个称之为“像素” (Pixel) 的规则小正方形格子组成的。每个像素各有一种颜色，用 R、G、B（红、绿、蓝）这 3 种基本色来合成，每种基本颜色分量都分成了 256 个等份 (0 ~ 255)。以 R（红色）为例，0 表示不含有任何红色，255 表示红色的含量是满的。如果，我们把左边的彩色图像转换成灰度，就变成了右边的黑白照片。实际上，用“黑白”这个词是不准确的，因为黑白只表示两种色彩：非黑即白。而灰度 (Gray) 实际上也分成了 256 个等分，表示 256 个等级的图像亮度，其中 0 表示全黑，255 表示全白。说起灰度，让我想起了网上的一句小诗：“如果生命的两端不是黑色的，人们又怎会爱上她灰色的中部？在生命灰暗的底色里，我愿做一条激情奔涌的河流，于波澜壮阔中探寻生命的灵动。”

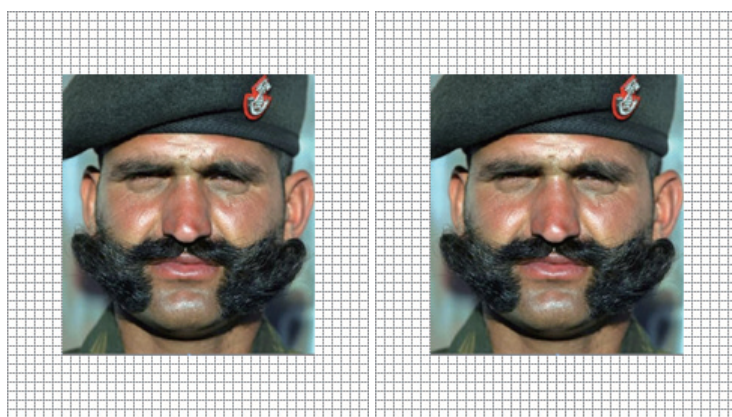


图 6-1 像素组成的彩色图像和灰度图像 (图片来源：Tiexue)

下面，以一种名叫 FAST (Features from Accelerated Segment Test) 的特征点检测方法为例，介绍一下特征的描述和检测。

FAST 特征点检测是公认的比较快速、简单、有效的特征点检测方法，只利用周围像素的比较信息就可以得到特征点。FAST 特征检测算法来源于 Corner (角点) 的定义：只选取候选点 p 周围的一些像素点进行比较，来确定此点是否是特征点。如果周围半径内有足够多的点都与候选点的像素灰度值差别很大，则认为该候选点是特征点。

这其实很好理解：一位美女长得好不好看，我们往往是通过把她和周围的其他女生进行比较来评价的。只要她能在这个局部小圈子中鹤立鸡群，她就是一个吸引眼球的“特征点”。如图 6-2 所示是一个半径为 3 的特征点的例子，因此周边共有 16 个像素点需要比较。

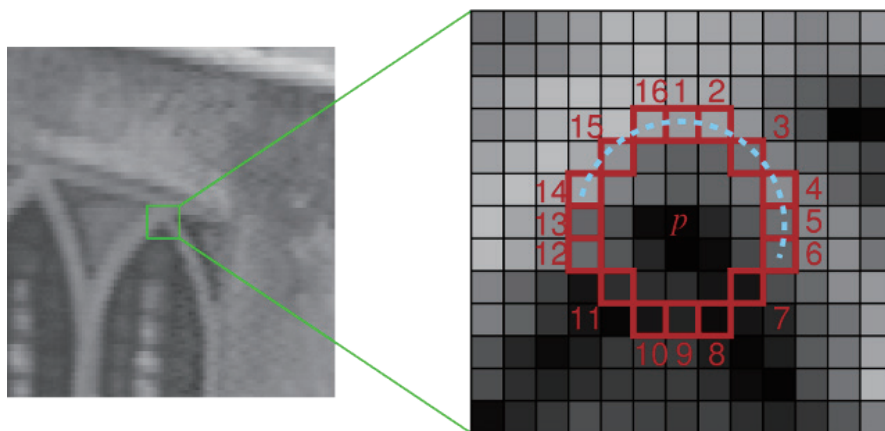


图 6-2 半径为 3 的特征点判定（周边共有 16 个像素点需要比较）（图片来源：Edward Rosten）

以上思路可用非常简单的符号来表示，一目了然。设 $g(p)$ 为处于中心的候选点 p 的灰度值， $g(q)$ 为 p 周围某一点 q 的灰度值，如果它们的灰度值差 $|g(p) - g(q)|$ 大于某个阈值 ε_d ，则可认为差别很大。我们数出与 p 点差别很大的周围点共有 N 个，若 N 大于给定阈值（一般为周围点数的四分之三¹），则认为 p 就是一个特征点。

在 OpenCV（见 6.11.1 节）中，非常简单的几条编程语句就可实现这个特征点检测算法。

```
// 定义所输出的特征点集合 keypoints
std::vector<KeyPoint> keypoints;
// 构造 FAST 特征描述符对象
FastFeatureDetector fast(40);
// 进行特征点检测
fast.detect(image, keypoints);
```

运行它！结果如图 6-3 所示。怎么样？确实很“智能”地把很多角点给找出来了，是不是非常简单？

OK，现在是不是信心满满？我们接着往下介绍一种名为 Haar-like 的特征。Haar-like 特征分为 4 类：“边缘特征”、“线特征”、“中心环绕特征”和“对角线特征”，如图 6-4 所示。每一类特征还分很多种，每种构成了一个特征模板。以边缘特征为例，共有正的、斜的 4 种特征模板。每个特征模板内有白色和黑色两种矩形，该模板的特征值定义为：“白色区域像素之和”减去“黑色区域像素之和”。

1 为了使速度更快，还可采用额外的加速。比如，测试候选点周围每隔 90° 角的 4 个点，应至少有 3 个和候选点的灰度值差足够大，否则就不用再计算其他（ $16 - 4 = 12$ ）点了，直接认为该候选点不是特征点。



图 6-3 FAST 特征点检测案例

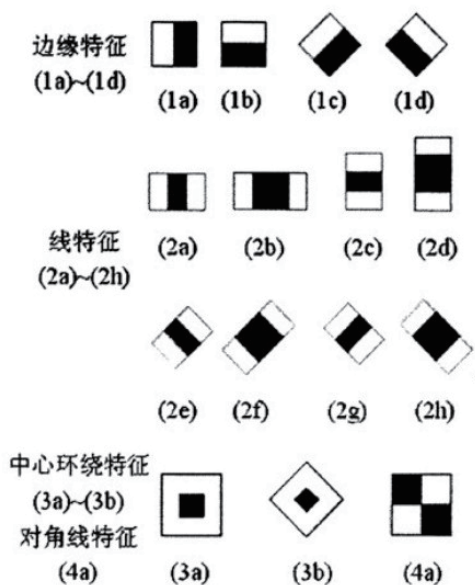


图 6-4 各个 Haar-like 特征模板

可能你会有些迷糊：这些所谓的特征不就是一堆堆带条纹的矩形吗？有正的，还有斜的，到底是干什么用的呢？我这样给出解释，将上面的任意一个矩形放到人脸区域上，然后，将白色区域的像素和减去黑色区域的像素和，得到的数值我们暂且称之为人脸特征值。比如将线特征的某个矩形模板放在眉毛处，得到了一个特征值 a 。但如果你把这个矩形模板放到一个非人脸区域，比如猴子的眉毛处，那么计算出的特征值 b 应该和人脸特征值 a 是不一样的。而且我们希望越不一样越好（也即可区分性越好），这些矩形方块的目的就是把人脸特征量化，以区分人脸和非人脸。

可以看出，Haar-like 是非常原始、粗粒度的，但优点是计算速度快。因此，为了增加区分度，可以对多个矩形特征组合（比如将检测眉毛的某个线特征模板和检测脸部轮廓的某个边缘特征模板进行组合）得到一个区分度更大的特征值。那么什么样的矩形特征怎样地组合到一块可以更好地区分出人脸和非人脸呢？这就是 AdaBoost 算法要做的事了，详见本章第 6.11.1 节“OpenCV 与 AdaBoost 人脸检测”。

最后再介绍一种直方图统计特征。**直方图（Histogram）** 又称柱状图，是一种统计报告图。灰度直方图是指对图像的灰度信息进行统计，它表示图像中每种灰度级（0 ~ 255）的像素的个数，反映图像中每种灰度出现的频率。

直方图是图像最基本的统计特征：直方图的横坐标是灰度级（0 ~ 255），纵坐标是该灰度级出现的频率（也即个数）。如图 6-5 右上所示，从这张灰度直方图（将彩色图像转换成灰度图像，再做直方图统计）可以看出，有 4 种灰度占了很大一部分比例，分别对应头发、衣服、皮肤及背景墙面。注意：不同的直方图代表不同的图像，同一直方图却未必表示图像相同。

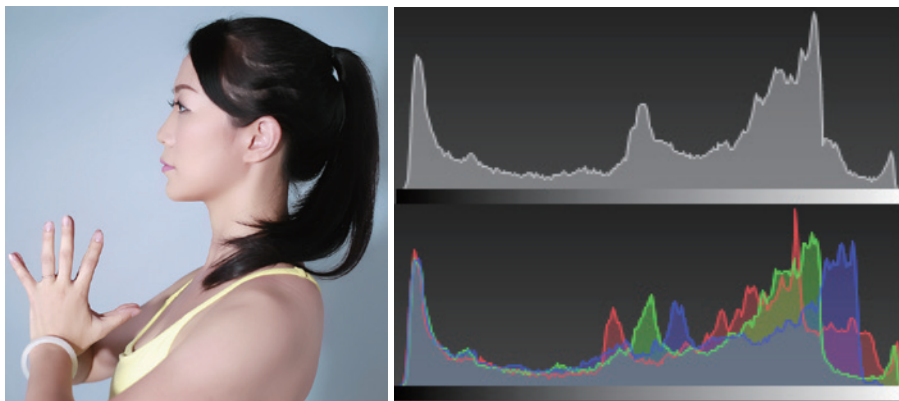


图 6-5 统计像素的灰度和颜色个数：灰度直方图（右上）和彩色直方图（右下）

除了灰度直方图，还经常直接用到彩色直方图（如图 6-5 右下所示），也即不用将彩色图像转换成灰度图像。最常用的颜色空间有：RGB 颜色空间、HSV 颜色空间。彩色直方图能够简单描述一幅图像中颜色的全局分布，即不同色彩在整幅图像中所占的比例，特别适用于描述那些难以自动分割的图像和不需要考虑物体空间位置的图像。其缺点在于：它无法描述图像中颜色的局部分布及每种色彩所处的空间位置，即无法描述图像中的某一具体的对象或物体。这也是统计本身的特点所决定的。比如，我只告诉你房间里有 4 个学生，那么这 4 个人都处在什么具体位置？身高、体重各为多少？有几个男生几个女生？是否戴眼镜？等情况就不得而知了。

小结一下。本节中共介绍了 3 种类型的特征，其中特征提取最重要的一个特性是“**可重复性**”：同一场景的不同图像所提取的特征应该是相同的。比如一位美丽的姑娘的右嘴角上有一颗美人痣，那么无论给她拍多少张照片，每张照片的右嘴角位置都应该有这颗美人痣。对于多视角特征匹配而言，可重复性约束是关键。

图像特征描述的核心问题是**不变性（健壮性）**和可区分性。不变性指的是对视角变化的不变性、对尺度变化的不变性、对旋转变化的不变性、对形状的不变性等，这对于物体识别中处理不同视角、遮挡、复杂背景等复杂情况尤为重要。

一个优秀的特征描述符还应该具有很强的**可区分性**。比如，一个好的特征描述符甚至可以区分出两个双胞胎姐妹外貌的细微差别。然而，特征描述符的可区分性的强弱往往是和不变性相矛盾的，也就是说，一个具有众多不变性（如旋转不变性、尺度大小不变性）的特征描述符，它区分局部图像内容的能力就稍弱；而一个非常容易区分不同局部图像内容的特征描述符，它的**健壮性**（不变性）往往比较低。如我们通过统计局部图像灰度直方图来进行特征描述，这种描述方式具有较强的不变性，对于局部图像内容发生旋转变化等情况比较**健壮**，但这种统计特征的区分能力较弱（或者说描述能力不够精确），无法区分两个灰度直方图相同但内容不同的局部图像块。更形象的例子：两张直方图都统计出房间里有 4 个学生，但一个房间是 4 个学生站在 4 个角落，另一个房间却是 4 个学生站在房间中间。

除了本节介绍的 3 个特征描述符，目前还有很多优秀的其他特征描述符，如 SIFT（Scale Invariant Feature Transform）、SURF（Speeded Up Robust Features）、DAISY、MROGH、BRIEF 等，各适用于不同的应用场景。上述这些特征描述符都是基于手动设计得到的，也有一些研究试图

利用机器学习的方法，通过数据驱动得到想要的特征描述符。这类特征描述符包括 PCA-SIFT、Linear Discriminative Embedding、LDA-Hash 等。特别是深度学习（见第 4 章 4.6.3 节“深度学习：像人脑一样深层次地思考”），使得自动学习特征成为了可能，而无须专家仔细设计和选择。

6.2.2 个性特征的定位与匹配

在上一节中我们通过特征描述符自动检测到了各种特征，如角点特征和边缘特征。但是，计算机是不知道哪个角点是眼角、哪个是嘴角、哪条线是眉毛的，这就需要进行特征的自动定位和匹配。国内外学者做了大量关于人脸特征点定位方面的研究。最初，一些学者利用视觉底层的特征，如上节中提到的角点、边缘提取等，但是这些方法的定位效果不够精确。后来，人们认识到仅用一些低级特征和基本的图像处理方法是远远不够的，应该对研究对象建立相应的先验模型（即预先定义的模型，比如我们告诉计算机两只眼睛共有 4 个眼角），以便进行精确的人脸定位。

我们先来看看如图 6-6 所示的各种手的形状，可以说各不相同，非常“个性化”。



图 6-6 各种手的形状

那么，我们能否提供一个统一的数学模型，对这些手的形状进行统一描述呢？这就需要我们首先把这些形状的“共性”提炼出来，然后加入一些个性化的参数信息，这就可以用这个先验模型来精确描述某一具体的形状了。比方说，我们要画一只手的形状，正常人的手指都是 5 个，这就是共性信息；而同时每个人的手指粗细长短各不相同，这就是个性信息；在共性的基础上加入这些个性化参数信息，就可以把某一个人的手精确勾画出来。

由于计算机直接表示复杂曲线的形状是比较困难的，所以我们可以用一定数量的**标记点（Landmark Points、Markers）**来近似。比如，我们可以在手的轮廓上标记很多个特征点来描述手的形状，如图 6-7 所示。这种模型就称为**点分布模型（PDM, Point Distribution Model）**。

本节中我们将重点介绍人脸特征定位，可广泛用于人脸识别、表情编辑、人脸跟踪、年龄估计、司机的驾车安全提醒等。这里介绍最著名的**主动外观模型（AAM, Active Appearance Models）**^[21]，其也是属于点分布模型的一种。AAM 首先采用统计分析的方法对几百幅手工标记的人脸图像进行训练，构建出人脸的形状模型和纹理（纹理可理解为图像的颜色或明暗变化信息）模型。根

据建立好的这个模型，不断地“主动（Active）”调节模型参数（包括形状参数和纹理参数）对图像中的人脸进行匹配运算。一旦匹配完成，也就同时实现了将 AAM 模型上预定义的那些点定位到图像上的人脸轮廓上。



图 6-7 手的特征点分布模型

下面我们对形状模型和纹理模型分别进行介绍。

形状模型

形状模型需要对训练数据（几百幅人脸图像）中每一幅图像手工标注特征点，以此来获取人脸样本的形状信息。比如，为了建立模型，我们对每一幅训练人脸标注了 68 个特征点，这些特征点标记在脸的外部轮廓和五官的边缘上，覆盖了人脸的整个形状，如图 6-8 所示。



图 6-8 标有特征点的训练人脸（图片来源：Tim Cootes）

这样，脸部的形状可以描述为一系列标记点的坐标集合，设标记点有 N 个（本例中 $N=68$ ），每个标记点由一个二维坐标确定（ X 轴和 Y 轴），因此所有这些标记点将组成一个 $2N$ 维的向量，将某一人脸样本的形状 \mathbf{s} 描述为一系列标定点的坐标集合：

$$\mathbf{s} = (x_1, y_1, \dots, x_n, y_n, \dots, x_N, y_N)$$

其中 \mathbf{s} 指某一人脸样本的形状, (x_n, y_n) 代表图像上第 n 个标记点的 X 坐标和 Y 坐标。

形状归一

由于样本库中每张人脸的大小、位置、摆放角度都存在很大的差异, 为了便于对库中所有人脸的形状进行统计建模, 应先对这些图像形状进行归一化。而归一化是指以某一个参考形状为基准, 将其他的形状进行旋转、平移和缩放并尽可能地与参考形状对齐。



注意：归一化只是对形状进行了对齐, 并没有对形状进行形变, 如非均匀拉伸和扭曲。

图 6-9 是几百幅人脸图像中所有特征点的集合, 以及把这些形状对齐后的集合。从左右两幅图的对比中可以看到, 对齐后的形状点基本在均值形状附近变化。这表明, 人脸形状确实可以用一个统一的模型来进行描述, 无论是描述共性, 还是个性化的差异。

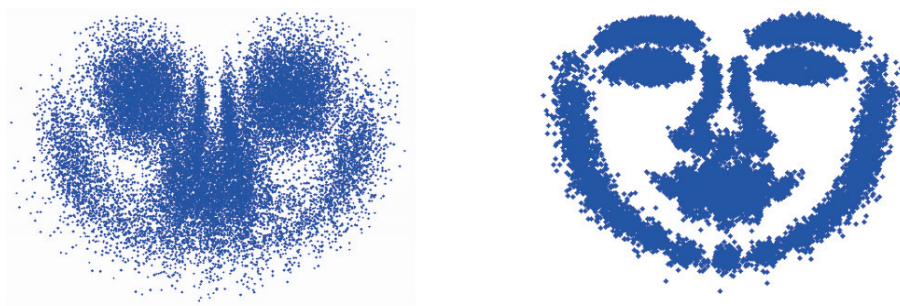


图 6-9 左：形状对齐前的点集合；右：形状对齐后的点集合

然后, 对齐后的数据进行**主成分分析 (PCA, Principal Component Analysis)**, 将这些原本海量高维的数据降到低维空间, 保留住信息量大、最能代表原事物的成分 (维度), 而将信息量小或冗余相关的成分 (维度) 去掉。这样, 任何人脸形状都可以用下面的公式来合成:

$$\mathbf{s} = \bar{\mathbf{s}} + \Phi_s \times \mathbf{b}_s \quad (1)$$

其中, $\bar{\mathbf{s}}$ 是对所有人脸形状进行平均后得到的平均形状 (常量), Φ_s 是 PCA 后所得到的人脸形状特征矩阵 (常量), \mathbf{b}_s 是形状模型参数 (变量)。这样, 如图 6-10 所示, 我们只需简单地改变参数 \mathbf{b}_s 值的大小, 就可以合成出不同的人脸形状 (其中, 正中间的人脸形状为“看上去完美无缺陷”的平均形状 $\bar{\mathbf{s}}$)。所谓合成人脸, 也即生成一张人脸库中原本不存在的人脸, 比如先取刘德华人脸形状的 70%, 再取赵本山人脸形状的 20%, 最后再取赵薇人脸形状的 10%, 合成得到一张全新的人脸。



提示：AAM 模型本质上是一种“形状混合人脸”技术, 其认为所有人脸构成了一个线性空间, 任意一张人脸可以用有限个人脸基的线性组合来逼近。



图 6-10 改变参数 b_s 来合成不同的人脸形状

纹理模型

纹理模型的构建方法与形状模型类似，首先把每个训练样本形变对齐到同一个参考形状，使两幅图像中人脸区域的像素点数目一致，且使特征点完全对齐（左眼角对左眼角、眉心对眉心等）。将人脸形状形变到参考形状的方法有很多类，有基于三角形网格的（如图 6-11 所示）、基于对应点的、基于线段的形变方法等。

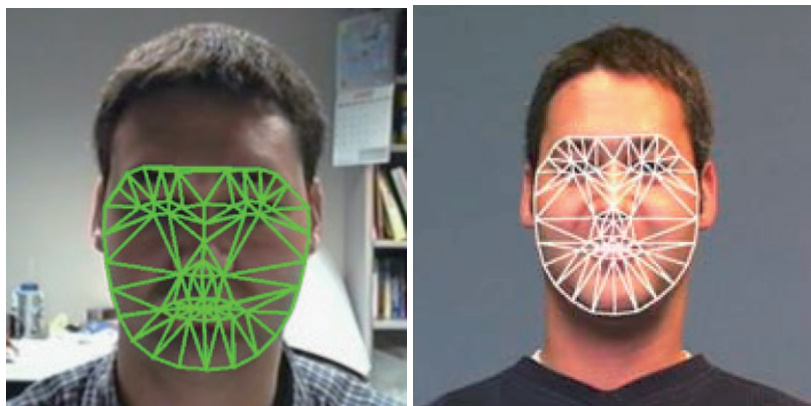


图 6-11 将两幅人脸形状用三角形网格进行对齐（图片来源：CMU）

然后从变形对齐后的图像中提取每个像素点的颜色灰度值，并把这些纹理信息（在 AAM 中，纹理即为图像灰度值）记为向量 \mathbf{g} （其中 M 为图像中人脸区域的像素点总个数）：

$$\mathbf{g} = (g_1, g_2, \dots, g_M)$$

对这些对齐后的纹理向量做 PCA 处理就可以得到纹理模型：

$$\mathbf{g} = \bar{\mathbf{g}} + \Phi_g \times \mathbf{b}_g \quad (2)$$

其中， $\bar{\mathbf{g}}$ 是对全部人脸样本进行平均后得到的平均纹理（常量）， Φ_g 是体现颜色灰度变化模式的特征矩阵（常量）， \mathbf{b}_g 是纹理模型参数（变量）。类似地，如图 6-12 所示，我们只需简单地改变参数 \mathbf{b}_g 的值，就可以合成出各种不同的人脸纹理！其中，中间的人脸纹理为“看上去完美无缺陷”的平均纹理 $\bar{\mathbf{g}}$ （是不是最红润光滑、吹弹可破？）。



图 6-12 合成的各种人脸纹理，其中中间为平均纹理

将上面的形状模型（1）和纹理模型（2）进行融合，我们就得到了主动外观模型（AAM），将其人脸形状和人脸纹理用一个统一的公式来表达：

$$a = \bar{a} + \Phi_a \times b_a$$

其中， \bar{a} 是对全部人脸样本进行平均后得到的平均外观，也就是第 5 章 5.3.3 节“人是种视觉动物：如何美化你的照片”中被视为美学一大发现的中性脸或者完美脸（不仅形状最完美而且肤色最佳）。这时，我们只需通过改变参数 b_a 的值，就可以用 AAM 合成各种不同的人脸图像 I_a 了。

使用 AAM 进行人脸特征点定位，实际上就是用 AAM 模型生成一张合成图像 I_a 去逼近目标人脸图像 I ，即最小化两者的差值 $(I_a - I) \rightarrow 0$ ，如图 6-13 所示。这个过程需要不断地调整 b_a 来生成最逼近的合成图像，其本质上是一个非线性迭代优化问题。目前有各种快速的求解方法，如反向组合图像对齐（ICIA, Inverse Compositional Image Alignment）算法等，可获得实时（最快可达 230 帧/秒）的人脸特征点定位和匹配。

图 6-13 左：目标人脸 I ；右：主动外观模型合成的人脸 I_a

如果我们将可变形模型的思想由二维图像扩展到三维，就有了**三维形变模型（3DMM, 3D Morphable Model）**^[26]，如图 6-14 所示。注意：这次最左边的是平均脸，而不是中间的。



图 6-14 从左到右：平均脸、50% 的目标人脸、目标人脸（图片来源：Volker Blanz）

特征匹配定位成功后，我们就自然实现了从单张正面照片进行 3D 人脸重建，如图 6-15 所示，参见第 5 章 5.3.1 节“基于单张照片的 3D 人脸重建及立体浮雕”。

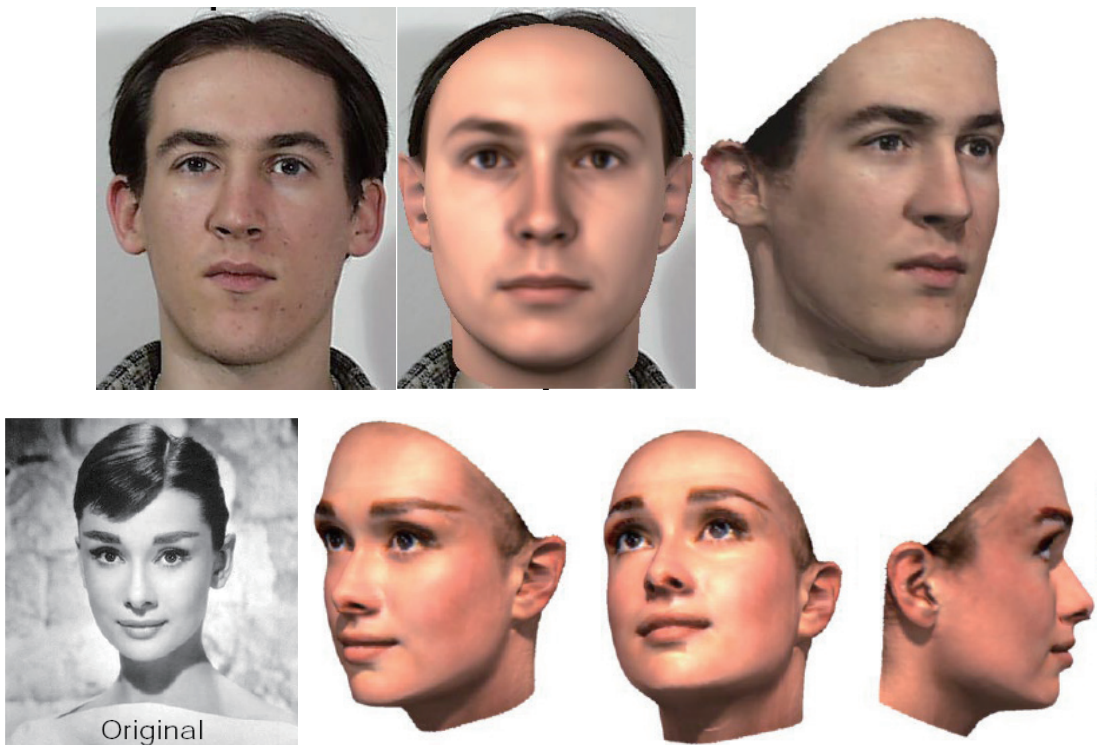


图 6-15 通过单张图片进行 3D 人脸重建，并将照片切换到新视角

6.2.3 个性化形状的编辑与合成

我们经常需要对形状做一定的编辑处理，比如将图 6-16 中左边的龙的嘴巴打开，生成右边的形状。

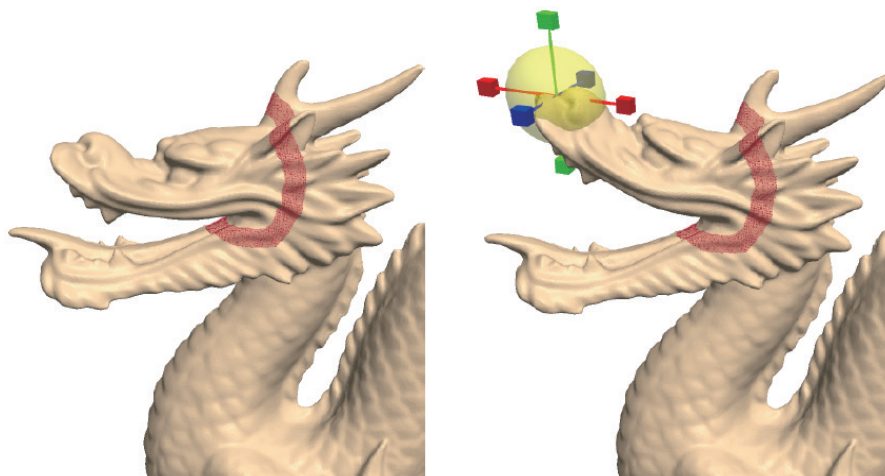


图 6-16 对龙的嘴部形状进行编辑（图片来源：Olga Sorkine）

这里涉及**细节保持 (Detail-Preserving)**的形变 (Deformation) 编辑 (Editing) 技术^[27]。如图 6-17 的左边所示，每个 3D 模型实际上由一个个顶点，以及连接它们的细小三角形面片连接而成。

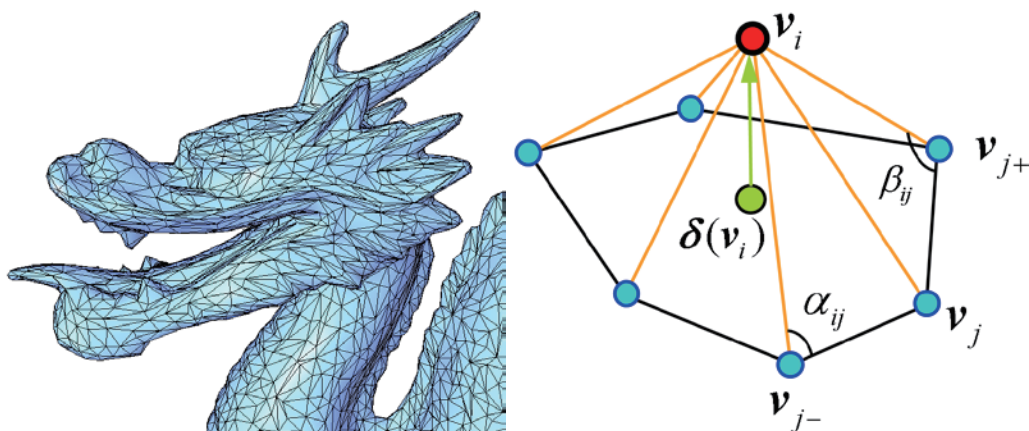


图 6-17 局部形状细节的编码

图 6-17 右边的图对某一个顶点 v_i 以及它周围的一圈顶点进行了放大显示。我们可对这个局部区域的形状（朝向、曲率²大小、局部面积、连接结构、夹角等）进行一个编码并保存起来，得到一个向量 $\delta(v_i)$ ：

$$\delta(v_i) = \sum_{j \in i^*} w_{ij} (v_i - v_j)$$

其中， i^* 是与顶点 v_i 相邻接的一圈顶点集合； w_{ij} 包含了顶点 v_i 与某一邻接顶点 v_j 的局部参数化信息，如连接拓扑结构、夹角 α_{ij} 和 β_{ij} 等。

不要小看了这个局部形状编码 $\delta(v_i)$ ，“牵一发而动全身”。因为所有的顶点都是相连的、“一

2 曲率可理解为中心顶点高出周围一圈邻接顶点的方向和程度，也就是局部的凹凸情况。

环紧扣一环”的，如果我们对所有顶点都计算各自的局部形状编码，并统一进行全局处理，那么就可对全局形状进行形变编辑和合成。

所谓细节保持的形变编辑，指的是在形变之后，总体的形状姿态虽然发生了改变，如图 6-16 所示的龙的上颚的朝向发生了改变，每个顶点的局部形状向量 $\delta(v_i)$ 的朝向也都发生了改变。但是，局部的细节却是刚性保持的。用公式表示就是，新点 v'_i 的局部形状向量 $\delta(v'_i)$ 应该是由原来点 v_i 的局部形状向量 $\delta(v_i)$ 经刚体变换 T_i 得到，这里我们采用最小二乘形式（假设 3D 模型共有 K 个顶点）：

$$\min f(V') = \sum_{i=1}^K \|T_i \delta(v_i) - \delta(v'_i)\|^2$$

具体来说：每个顶点的局部形状向量 $\delta(v_i)$ 的长度 $\|\delta(v_i)\|$ （即该点的曲率大小）应该是保持不变的，而且公式中的局部结构参数 w_{ij} （即与周围点的连接结构、夹角）也应该是保持不变的。唯有这样，形变编辑后的龙的上颚才能看起来是龙的上颚形状，而不是一只老鼠的上颚形状，抑或是被揉成了一团面糊的杂乱形状。更多、更具挑战性的大角度形变如图 6-18 所示。



图 6-18 大角度形变的例子



提示：三维空间中的形变，在数学本质上是一个非线性系统。小角度的形变因为变化不大，所以可用一个线性系统来近似地快速求解。而大角度形变，要么直接费时地求解非线性系统，要么通过迭代多次线性系统来逼近，还可由粗到精的多尺度策略来分而治之。

基于局部形状编码 $\delta(v_i)$ （其实它还有一个很学术的专业名称：**拉普拉斯微分坐标 Laplacian Differential Coordinates**），我们还可实现一种很有趣的操作：**形状合成（混搭，Shape Composition、Shape Mixing）**。如图 6-19 所示，我们可以将一堆形状部件“粘”在一起，而在部件之间的交界处，形状的过渡会保持平滑、自然——这可通过对交界处的形状微分坐标进行插值并把产生的局部扭曲失真均匀地扩散分摊到整片区域来获得。



扩展：形变方法主要分为基于表面的和基于空间的方法。**基于空间的形变方法**不直接形变物体，而是通过对物体所嵌入的空间进行形变来对物体形状进行修改；类似于你弯曲一块汉堡，则夹在里面的热狗也会跟着变形。代表性的方法有**自由形变（FFD）**方法，其可分为基于栅格的，基于曲线的，或基于点的方法。同时还可定义不同的基函数来控制空间形变，如基于径向基函数（Radial Basis Function）或涡状（Swirl）函数。在处理空间形变时需要考虑的问题是避免自相交，以及保持全局或局部体积。

而本节中所介绍的**基于表面的形变方法**直接在形状的表面定义形变。作为一种局部内蕴特征描述符，**离散微分属性（如拉普拉斯坐标或者梯度场）**已经被用于细节保持的网格编辑。然而，顶点的拉普拉斯微分坐标是个带方向的向量，因此不是旋转不变的，须通过某种方式进行转换以匹配到期望的新方向，如显式地通过启发式的方法（或用户调整）或隐式地采用迭代非线性方法，否则网格细节形变后将产生扭曲。

图 6-19 分别是美国 UIUC 大学、微软亚洲研究院、浙江大学（茶壶与门环铺首^[28]）、以色列 Tel Aviv 大学（兔子与翅膀^[27]），以及 Autodesk 的 MeshMixer 进行形状合成的例子，是不是很魔幻、很“九头怪”化？利用形状合成，你完全可以先从网上找到“钢铁侠”的 3D 全身模型，然后只将他的脸部替换成你的脸，相当于你就穿上“钢铁侠”的行头了！然后，将这个“**换脸**”后的全身模型 3D 打印出来，摆在自己的桌上。

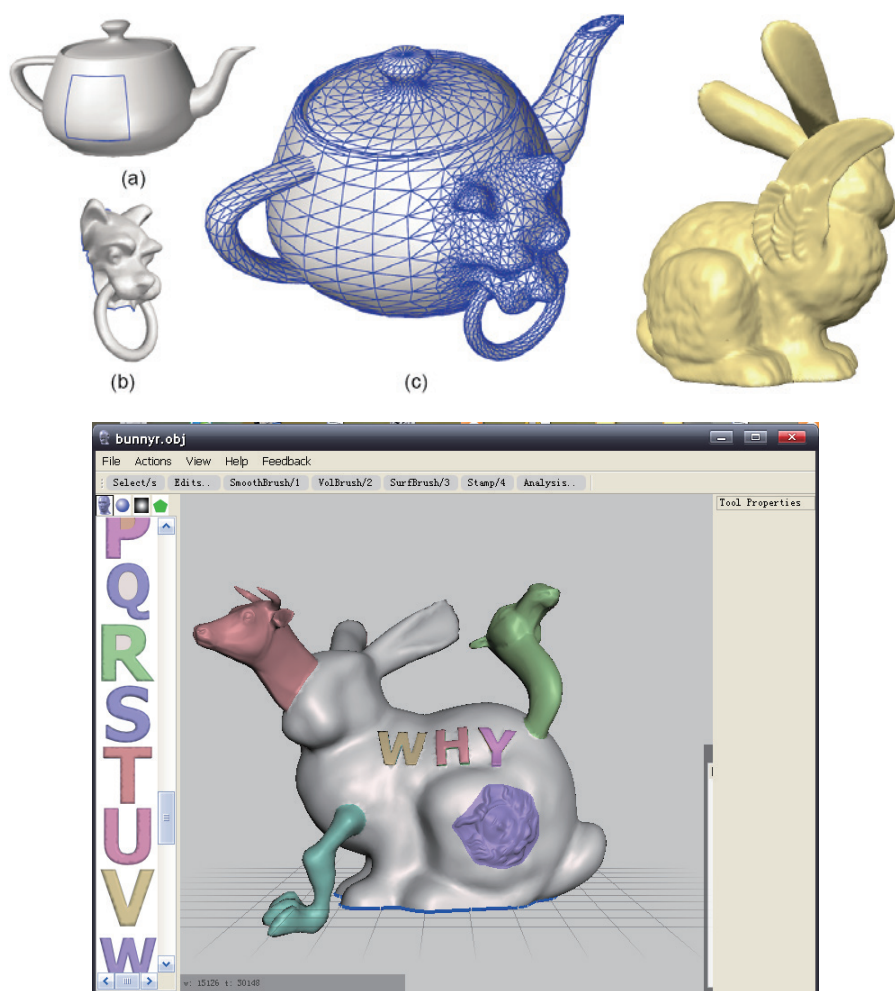


图 6-19 形状合成的例子（图片来源：UIUC、微软、浙江大学、Tel Aviv、Autodesk）

形状合成将一个模型（大多数情况下只是“残肢、断腿、换脸”）嫁接到另一个模型上，还有一种更酷的技术叫**形状混合（Shape Blending）**、**形状插值（Shape Interpolation）**，它将两个（或

多个)模型进行逐点整体融合。如图 6-20 上方所示,我们将一只狮子的 50%,一只猫的 30%,一只骆驼的 20% 合在一起,进行“混血杂交”,得到图 6-20 上方右边的新物种模型。是不是有点意思呢?如图 6-20 下方所示,我们将一位少女**形状渐变 (Morphing)**成一只麋鹿,中间生成了一系列的模型,代表着不同比例的形状混合效果。形状渐变在电影动画中应用很多,比如变形金刚中的汽车变机器人。

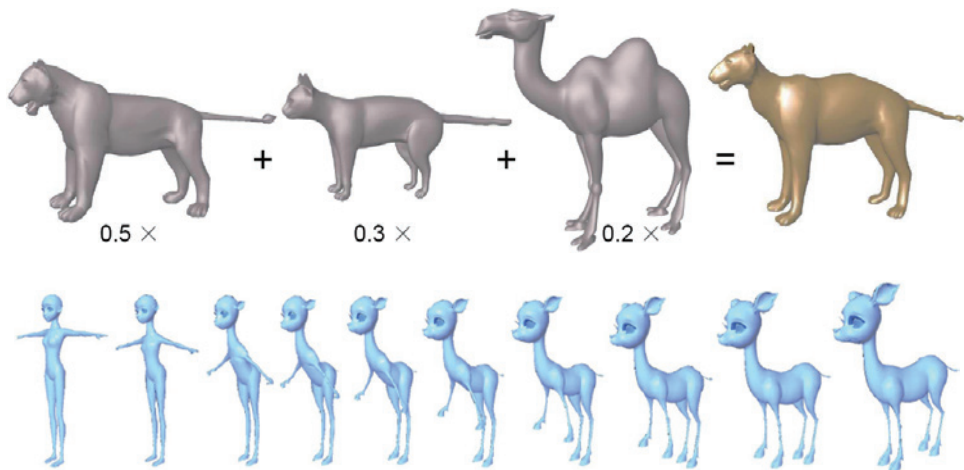


图 6-20 形状混合与形状渐变

要实现形状渐变,首先必须在两个(或多个)模型之间建立一对一的双射(Bijective Mapping)。这个过程也被称为**一致对应 (Consistent Correspondence)**或**交叉参数化 (Cross-Parameterization)**,即,将具有不同拓扑连接的 3D 模型(比如狮子模型 S 有 5 000 个顶点,骆驼模型 T 有 10 000 个顶点)转化为**兼容网格 T' (Compatible Mesh)**。兼容网格指具有相同拓扑的网格(比如让骆驼模型变形成狮子模型的形状,但仍保持 10 000 个顶点),同时又保持住细节特征(不能有大的失真扭曲,比如骆驼眼睛本应变形成狮子眼睛圆圆的形状,却弄成三角眼了;骆驼表面原本均匀的局部参数化信息,你可想象成骆驼腿上穿了带有均匀条纹的黑色长丝袜,被扭曲得错乱畸形,甚至翻转褶皱了),其中涉及**保角映射 (Conformal Mappings)**。在建立一致对应之后,剩下的事情就相对简单了,确定对应点的**插值路径 (Trajectory)**即可。最简单的方法是直接对模型 T 和 T' 线性插值,当然,你还可使用非线性插值(如非线性梯度场插值)以克服大角度旋转可能引发的塌缩(Shrinkage)现象。



提示:除了直接构建两个 3D 模型(S 、 T)的一致对应,还可采用**间接**的方法,即构造一个中转用的公共参数化域 C (如平面、球面、柱面、三角分区平面等),之所以选择这些 2D 凸平面域或球域,是为了便于建立保角映射。然后,两个模型就被分别映射到这同一个公共参数化域上去,并获得了两个保角子映射(φ_{SC} 、 φ_{TC})。最后,两个模型的一致对应(映射)就可通过两个子映射的合成来获得($\varphi = \varphi_{TC}^{-1} \circ \varphi_{SC}$)。这其实很好理解,为了建立一条大胖鱼和一条小瘦鱼全身的一一对应,我们把它们缩小或放大对齐到同一大小的中间域即可,这样它们全身每个点就可以很自然地被一一对应起来。间接方法的难点在于如何高效并健壮地对复杂形状构造良好的(well-shaped)兼容分区(compatible patches),尤其在拓扑复杂或者狭长的区域时。

本节介绍的只是简单的几何形变编辑。如果需要进行更复杂的高层语义编辑，如改变人脸表情（让3D人脸模型对你呵呵地傻笑），请参考第5章5.6节“3D人脸表情形变与编辑”。

6.3 立体视觉重建：将照片转成3D数字模型

我们之前提到过，为了实现3D打印的定制化，很多情况下需要用到计算机视觉技术。比如要为自己量脚定制一双鞋，用户拿手机对着自己的脚拍下多个角度的照片，单击上传按钮，云端的智能计算服务就可以自动把3D脚型重建出来，然后设计出匹配的鞋子并将其3D打印出来。下面我们就以立体视觉重建为例，介绍一下计算机视觉技术。本节是第5章5.3.2节“基于多视角照片的3D人脸重建”的原理基础。

6.3.1 摄像机定标

假设你现在已经拍摄了脚的多张各个角度的2D照片，那么如何将这些照片转化成一个3D数字化形状呢？首先第一步，你要对摄像机进行定标，比如确定摄像机的焦距、摆放位置和角度等。

我们先来看看摄像机的成像模型^[1]。摄像机的成像模型一般采用**针孔模型（Pin-Hole Model）**，与我们中学时学的小孔成像是一个原理，如图6-21所示。

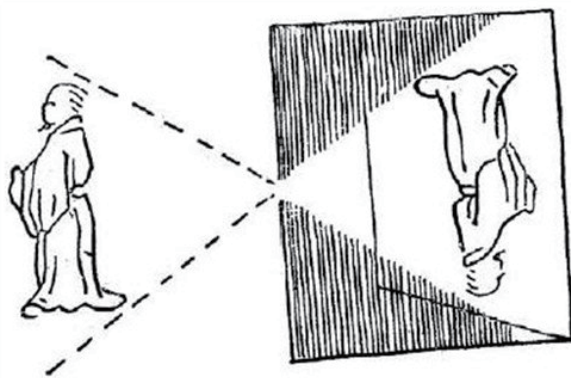


图6-21 两千多年前，墨子和学生进行了世界上第一个小孔成像实验（图片来源：大科技）



提示：其实两千多年前我国古代的大思想家墨子就和学生进行了世界上第一个小孔成像实验，并记录到“景到（即，倒），在午（即，交叉）有端（即，小孔），与景长，说在端”。因此墨子被西方称为“摄影光学理论和实践的开创者，探索光影成像的第一人”。

如图6-22左边所示，为了把成像模型解释清楚，我们来仔细看看摄像机的成像几何关系。同时我们把成像平面放到了小孔的前面，这样成像就是正立着的而不像上图那样倒立了。

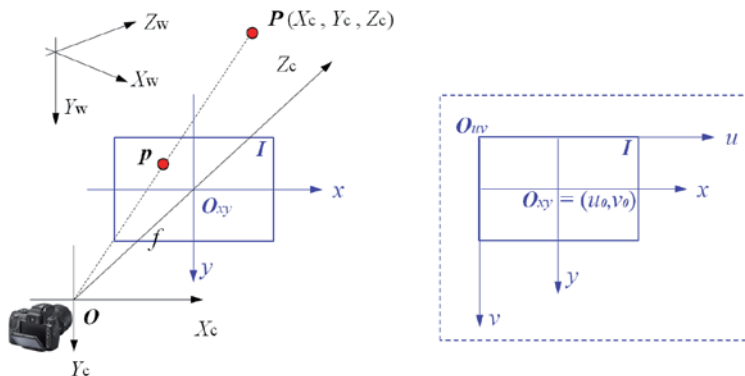


图 6-22 左：针孔成像模型；右：图像坐标系

- O 点称为摄像机的**光心**，由点 O 与 X_c 、 Y_c 、 Z_c 轴组成的直角坐标系称为**摄像机坐标系**。
- I 是成像平面（**图像平面**），我们把镜头对焦后，物体就成像在这个平面。图像平面构成了一个**图像坐标系**，横坐标为 x ，纵坐标为 y 。
- X_c 轴和 Y_c 轴与图像的 x 轴与 y 轴平行， Z_c 轴为摄像机的**光轴**，它与图像平面垂直。光轴与图像平面的交点，即为**图像坐标系的原点** O_{xy} 。
- OO_{xy} 的长度为**摄像机焦距** f 。

一下子冒出这么多术语，我们中场休息一下，给出本节的第 1 个公式来醒醒脑子。我们先仔细研究一下图像坐标系。如图 6-22 右边所示，图像坐标系以 O_{xy} 为原点，由 x 、 y 轴组成，单位是 mm。然而，在实际的相机中，并不是以物理单位（如 mm）来表示某个成像点的位置的，而是用像素的索引。比如一台相机的像素 \times 是 $1\,600 \times 1\,200$ ，说明图像传感器（也就是以前的胶片）横向有 1 600 个捕捉点，纵向有 1 200 个，合计 192 万个。对于某个成像点，实际上都是这样表示的：横坐标第 u 个点，纵坐标第 v 个点（而不是横坐标 x mm，纵坐标 y mm）。假设 O_{xy} 在 u 、 v 坐标系中的坐标为 (u_0, v_0) ，每一个像素在 x 轴与 y 轴方向上的物理尺寸为宽 dx mm，高 dy mm，则图像中任意一个像素的索引坐标与物理坐标满足下面的换算关系：

$$u = \frac{x}{d_x} + u_0$$

$$v = \frac{y}{d_y} + v_0$$

上面的这个方程组，分成了单独两个公式来写，让人感觉关联不够紧密（“本是一家人却说两家话”），因为实际上是要同时满足的，因此我们把两行合并成一个公式来写：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3)$$

上面的形式被称作为**矩阵**形式。形象地说，矩阵（Matrix，没错，电影《黑客帝国》的英文片名就是这个单词）就是将多个有关联的元素用个大括号括在一起成为一个整体，以便直接对整体进行运算以分析出各个元素之间的内在关系。一个 $m \times n$ 的矩阵是一个由 m 行 n 列元素排列成的矩形阵列；当 $m=1$ 或 $n=1$ 时，则特称为**向量**。上式左边是个 3×1 的向量，右边表示了一个 3×3 矩阵和一个 3×1 向量之间的相乘。改成矩阵写法之后，将两个公式浓缩成了一个公式，整体感是不是提升了很多？从上面也可以看出，与**普通代数**（传统代数）不同，矩阵代数这种**超凡代数**描述的是**多维**而不仅是一维的代数；普通（一维）代数的乘法与乘子顺序无关， $a \times b = b \times a$ ，而矩阵代数通常情况下 $A \times B \neq B \times A$ ，即一般不满足交换律；此外普通代数只有 0 没有逆（倒数），而矩阵代数中，很多矩阵都没有逆。



提示：有的读者可能会问，上面的矩阵为什么有 3 行呢？之前的公式不是只有 2 个吗？好问题！这是我偷偷塞了点料，把之前的坐标 (u, v) 和 (x, y) 都转成齐次坐标 $(u, v, 1)$ 和 $(x, y, 1)$ 了。**齐次坐标（Homogeneous Coordinate）**的好处是：即使乘个系数 k ($k \neq 0$)，仍对应于原来的同一个点，也即 $(u, v, 1)$ 与 $(k \times u, k \times v, k)$ 对应于同一个点。同时，还便于几何变换（旋转、缩放、平移），只需用一个大一号的矩阵即可将变换矩阵的乘法（旋转、缩放）和加法（平移）合并到一块。此外，齐次坐标还可表示不同的无穷远点。

好，我们接着进行术语教学。之前我们定义了摄像机坐标系，但由于摄像机可安放在环境中的任何位置，所以我们还要定义一个**世界坐标系**来描述摄像机在环境中的位置，由 X_w 、 Y_w 、 Z_w 轴组成。摄像机坐标系与世界坐标系之间的关系可以用旋转矩阵 R 与平移向量 t 来描述，即：

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = M_b \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (4)$$

矩阵和向量用**黑斜体**表示，其中 R 为 3×3 的矩阵； t 为 3×1 的向量； M_b 为 4×4 的矩阵，其也被称为**摄像机外部参数矩阵**。



知识点：你在高中学习的初等数学是研究常量、研究静态的数学；而在大学里学的高等数学是研究变量、研究运动的数学。下面我们再解释一下空间、向量、矩阵的关系。

空间（Space）是高等数学的基础概念，可被通俗地解释为“存在一个集合，在这个集合上定义某某概念，然后满足某些性质”。从拓扑空间开始，一步步可往上“升级”成更高级的空间。**线性空间**（又名**向量空间，Vector Space**）其实还是比较初级的，如果在里面定义了范数（“长度”），就成了**赋范（Normed）**线性空间。赋范线性空间满足完备性，就成了**巴那赫（Banach）**空间；若定义了角度，则为**内积（Inner Product）**空间。内积空间若同时满足完备性，就得到**希尔伯特（Hilbert）**空间（见下面的 6.4.2 节）。

以我们日常生活在其中的三维**欧几里得（Euclidean）**空间为例：1）由无穷多个位置点组成，2）这些点之间存在相对的关系，3）可在空间中定义长度、角度，4）**这个**

空间可以容纳运动。这里我们所说的运动是从一个点到另一个点的“跃迁”跳跃式地运动（**变换，Transform**），而不是微积分意义上的“连续”性的运动。实际上，不管什么空间，都必须容纳和支持在其中发生的符合规则的运动（变换）。比如拓扑空间中有拓扑变换，线性空间中有线性变换，仿射空间中有仿射变换，其实这些变换都只不过是应对空间中允许的运动形式而已。

在线性空间中选定基（坐标系）之后，**向量刻画对象，矩阵刻画对象的运动**，也即用矩阵与向量的乘法来施加运动。**矩阵的本质是运动（变换）的描述**，比如在一个线性空间中，只要我们选定一组基，那么对于任何一个线性变换，都能够用一个确定的矩阵来加以描述。

若矩阵 A 与 B 是同一个线性变换的两个不同的描述（之所以会不同，是因为选定了不同的基，也就是选定了不同的坐标系），则一定能找到一个非奇异矩阵 T ，使得 A 、 B 之间满足这样的关系： $T^{-1}AT=B$ 。这就是**相似矩阵**的定义，即这两个相似矩阵 A 与 B 实际上描述的是同一个线性变换，因此**特征值（Eigenvalue）**相同。

从另一个角度来看，矩阵是由一组向量组成的，如果矩阵**非奇异**的话，那么这一组向量是**线性无关**的，于是它们组成了度量线性空间的一个坐标系。换言之，矩阵实际上描述了一个坐标系。之所以矩阵既是运动，又是坐标系，是因为对象的变换等价于坐标系的变换。

此外， n 阶矩阵的**行列式**也有着明确的几何意义：为 n 个组成向量按照平行四边形法则所张成的一个 n 维立方体的体积（如果是 2 维，则为面积）。向量的**线性相关性**实际上表示了这些向量所张成的广义平行四边形面积（体积）为 0，例如此时的向量共线（ $n=2$ 情况下）或共面（ $n=3$ 情况下）。反之，若线性无关，则体积（行列式）不为 0。此外，我们也不难看出，行列式不为 0 的矩阵，是**可逆**的（非奇异），即可将一组线性无关的向量变换成另一组也保持无关性的向量。

如果一个 n 阶矩阵虽不能保持 n 个向量的线性无关性，但它能保持 $r < n$ 个向量的线性无关性， r 就被称作矩阵的**秩**，表示了能保持非 0 体积的几何体的最大维数。

外围的知识都介绍完了，好，我们现在正式开始介绍针孔模型。如图 6-22 左边所示，空间上任何一点 P 在图像上的投影位置 p 为光心 O 与 P 点的连线 OP 与图像平面的交点，这种关系也被称为**中心射影**或**透视投影**。由几何比例关系可轻易得出（先看下面公式的左边）：

$$\begin{aligned} x &= \frac{fX_c}{Z_c} \\ y &= \frac{fY_c}{Z_c} \end{aligned} \Leftrightarrow Z_c \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (5)$$

其中 (x, y) 为 p 的图像坐标， (X_c, Y_c, Z_c) 为空间点 P 在摄像机坐标系下的坐标。同样，上式的箭头右边为左边的矩阵形式，左右两边是等价的。

我们将公式（3）和公式（4）代入公式（5），就可以得到 P 点在世界坐标系下的坐标 (X_w, Y_w, Z_w) 与其在图像平面的投影点 p 的坐标 (u, v) 的关系：

$$\begin{aligned}
Z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \begin{pmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \mathbf{M}_a \mathbf{M}_b \mathbf{p}_w = \mathbf{M} \mathbf{p}_w
\end{aligned}$$

其中, $\alpha_x = f/d_x$, $\alpha_y = f/d_y$; $\mathbf{M} = \mathbf{M}_a \mathbf{M}_b$ 为 3×4 矩阵, 称为**投影矩阵**; \mathbf{M}_a 完全由 α_x 、 α_y 、 μ_0 、 ν_0 决定的摄像机内部结构 (如焦距、光心) 有关, 称为**摄像机内部参数**; \mathbf{M}_b 完全由摄像机相对于世界坐标系的方位 (如摆放位置和拍摄角度) 决定, 称为**摄像机外部参数**。确定某一摄像机的内部和外部参数, 就被称为**摄像机定标 / 标定 (Calibration)**。注意, 很多情况下的摄像机定标**仅指**确定摄像机的内部参数。

由上式可以看出, 如果已知摄像机的内参数 \mathbf{M}_a 和外参数 \mathbf{M}_b (即定标好了), 两者直接一相乘就可得到投影矩阵 \mathbf{M} 。投影矩阵 \mathbf{M} 有什么作用呢? 对任意空间点 \mathbf{P} , 如果已知它的世界坐标系坐标 $\mathbf{p}_w = (X_w, Y_w, Z_w, 1)^T$, 根据 \mathbf{M} , 就可求出它的图像点 \mathbf{p} 的位置 (u, v) 。

单台摄像机的定标

下面介绍一下摄像机定标的过程, 如图 6-23 左边所示, 我们在摄像机前放一个标定块。所谓**标定块**, 就是每个特征点的空间位置都被事先测定的基准块, 比如两个点的距离都被精确做成了 10 000 cm (精确到小数点后若干位)。用相机拍摄标定块的图像, 就可以根据特征点的图像坐标 (u, v) 与真实三维空间坐标 (X_w, Y_w, Z_w) 之间的关系, 来计算摄像机的内外参数了。我们需要 6 个或以上特征点, 就可求解出投影矩阵 \mathbf{M} 。但在实际应用中, 我们通常使用了几十个特征点以减少可能的误差。此外, 我们常用图 6-23 右边的**平面标定块**来进行标定^[71], 标定时先平放着拍一张正面照片, 然后再用个小物件轮流把每边往空间上翘起来一点 (即改变标定板的空间位置姿态), 依次拍摄 3 幅以上照片即可。(不要小看了小小的标定板, 比如一块选材做工精良的大理石标定板定价需要几千元, 以保证工业级高精度、完全平面、不变形。选购时, 比如要实现 $\pm 10 \mu\text{m}$ 的 3D 重建精度, 可选择 $1 \mu\text{m}$ 精度的标定板。)

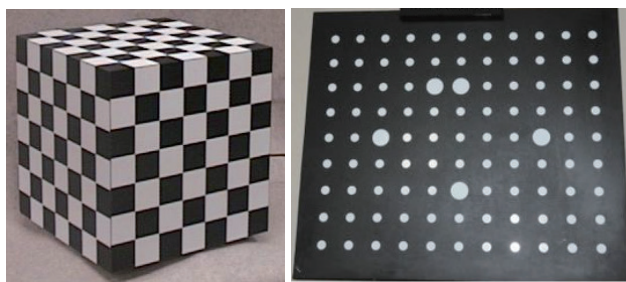
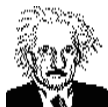


图 6-23 左: 三维标定块; 右: 二维标定板



提示：如果在定标时，环境不允许放置标定块或标定板，或者经常需要改变摄像机的内参数（如调焦），这时就只能通过**自标定（Self-Calibration）**技术来获取摄像机内参数，能利用到的信息只有图像的空间对应点，并用数学上的**绝对二次曲线（AC, Absolute Conic）**作为虚拟标定物。绝对二次曲线是在射影空间中无穷远平面上、全部由虚点构成的一条二次曲线，它的重要特性是它在图像平面的成像（IAC）不随摄像机的位置姿态变化，即投影成像只与摄像机的内参数有关。自标定至少需要拍摄3幅图像。另外对于平面场景的自标定，可通过提取平行线段进行求解。由于自标定求解的非线性，所以一般适用于精度要求不高的场合。

如果摄像机内参数 M_a 已知，仅需要确定外参数（ R 、 t ），即由 N 个 3D 空间点与图像点的对应关系来确定世界坐标系与摄像机坐标系之间的欧氏变换，这个问题被称为 **PNP 问题（Perspective-N-Points）**。当 $N \geq 6$ 时，可线性唯一确定；但 $N=3, 4, 5$ 时，一般情况下也可把解限定在有限的几个候选上。

以上我们假设的都是线性摄像机模型，但如果使用广角镜头（或全向摄像机），在远离图像中心处会有较大的畸变，因此可能还需要进行非线性修正或建模，如对径向畸变、离心修正、薄棱镜畸变等非线性畸变进行修正。

两台摄像机的定标

好了，我们刚才对一台摄像机的内外参数进行了标定。然而，对于从 2D 图像到 3D 形状的重建，一台摄像机（或仅拍一个角度的照片）是不够的。从图 6-22 中可以看出，已知图像点 p 的位置 (u, v) ，即使知道摄像机的内外参数，空间坐标 (X_w, Y_w, Z_w) 也不是能唯一确定的，实际上任何位于射线 OP 上的空间点的图像点都是 p 点。怎么办？再加一台摄像机（或者用同一台摄像机拍两次）！用左右两个摄像机进行**立体视觉**重建。

我们分别对左右两台摄像机单独进行了定标，就得到了它们的内参数。我们在 6.3.2 节可以知道，在一般的立体视觉方法中，只需知道摄像机的内参数，以及用**极线（Epipolar Line）**描述的双摄像机相对位置就足够了。

下面我们对极线进行介绍。如图 6-24 所示，用两个摄像机同时获得两张图像 I_1 与 I_2 。如果 p_1 和 p_2 是空间同一点 P 在两张图像上的投影点，我们称 p_2 为 p_1 的对应点，反之亦然。我们指出， p_1 点的对应点 p_2 不需要在 I_2 整幅图像中搜索，它必然只位于 I_2 的某一条直线上，该直线（图中的 l_2 ）称为图像 I_2 上对应于 p_1 点的**极线**。

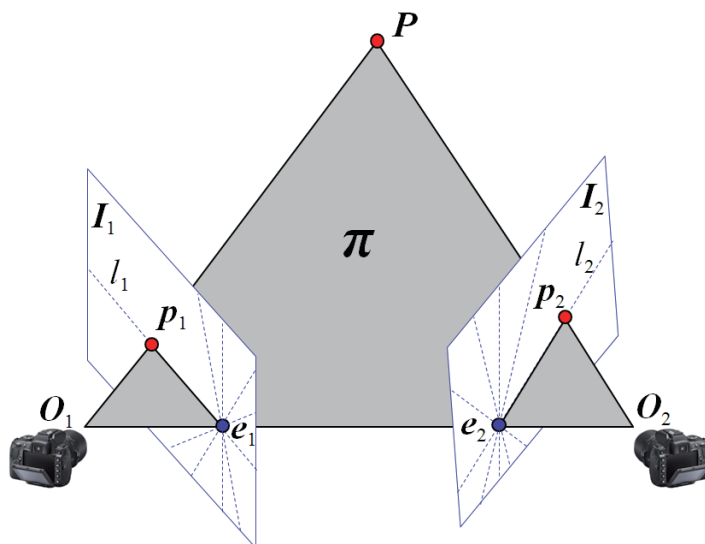


图 6-24 用两台摄像机拍摄空间点

那么如何找到极线呢？其实很简单，它满足一个几何约束。如图 6-24 所示， O_1 为左摄像机的光心， O_2 为右摄像机的光心，两者的连线 O_1O_2 经过左右两个图像平面时相交于两个极点 e_1 和 e_2 。可以证明， I_1 （或 I_2 ）图像上的所有极线都相交于同一点 e_1 （或 e_2 ），也即极线必然经过 e_1 （或 e_2 ）。经推导可知，给定投影矩阵 M_1 与 M_2 ，对于左摄影机的一个图像投影点坐标 p_1 ，则它的对应点坐标 p_2 所在的极线是完全确定的，满足如下方程：

$$p_2^T F p_1 = 0 \quad (6)$$

其中， $F = M_{a2}^{-T} [t]_x R M_{a1}^{-1}$ ；反对称矩阵 $[t]_x = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$ ，其秩为 2。注意 F

是一个仅为 3×3 的矩阵，具有 7 个自由度，由投影矩阵 M_1 与 M_2 组合运算得到，这是立体视觉很重要的一个矩阵，称为**基本矩阵（Fundamental Matrix）**，其只与两个摄像机的内参数和空间相对位姿有关，与外参数无关。

通过上式（6）还可以看出，无须费劲地求解出两个摄像机的所有内外参数（即 M_1 和 M_2 ），很多应用场合下这是不必要的，因为只需一个基本矩阵 F 就可以得到立体视觉的极线几何约束关系。为了求解出基本矩阵 F ，我们至少需要两幅图像中的 7 或 8 个对应点。

更广义地，我们介绍图像平面间的**单应矩阵（Homography Matrix） H** 的概念，具有 8 个自由度（实际上，前面提到的 F 由平面单应矩阵与极点唯一确定，即 $F = H^{-T} [e]_x$ ）。所谓**单应**可理解为：空间平面在两个摄像机各自的**射影变换（Projective Transformation，也被称为直射变换（Collineatory Transformation）**下所生成的图像点具有**一一对应**的关系，即 $p_2 = H p_1$ 。单应矩阵 H 的一个典型应用就是对射影变换导致的图像变形进行矫正：我们只需先从变形图像中指定共面的 4 个点，对其手工矫正后得到 4 个新的点，于是就可求解出单应矩阵 H 了，最后对变形图像的所有点应用单应变换 H 进行自动矫正，整个过程无须求解摄像机的任何参数。



提示：空间点 P 不仅可以通过在左右两张图像上的投影点 p_1 和 p_2 来重建（这称之为**双视图几何**），类似地，还可通过同时拍 3 张图像来重建点 P ，这就被称为三视图几何。其中三焦张量（Trifocal Tensor）等同于双视图几何中基本矩阵的地位，类推还有四视图几何的四焦张量。3 视图及以上的几何重建统称为**多视图立体重建（MVS, Multi-View Stereo）**。

6.3.2 基于立体视觉、SFM 和 Visual Hull 的三维重建

我们先介绍一下 3D 重建的基本原理。如图 6-24 所示，如果我们知道了 p_1 点以及它的对应点 p_2 的位置，并对两个摄像机进行了定标（得到两个投影矩阵 M_1 与 M_2 ，也即知道了焦距、光心距离等信息），那么很简单，可以用三角测距（见图 6-24 中灰色的大三角形 O_1O_2P ，其所在平面 π 因为经过了两个光心 O_1O_2 故也被称为**极平面**）的原理来计算空间点 P 的三维坐标。我们将这个过程称之为基于立体视觉的三维重建。两个摄像机光心的连线 O_1O_2 也被称为**基线（Baseline）**，即两个摄像机的光心距离。基线越长，则重建误差越小，但基线长度不可太长，否则不仅会造成点的匹配困难，而且可能由于物体各个部分的相互遮挡，导致两个摄像机无法同时观察到空间点 P 。当两个摄像机仅存在水平方向上的纯平移运动时， p_1 与 p_2 的纵坐标相同， p_1 与 p_2 横坐标之差 $u_1 - u_2$ 称为**视差（Disparity）**，这也是我们看 3D 立体电影觉得真实的原因：左右两只眼睛看物体会位置上的细微偏移。空间点 P 离摄像机越远，则视差越小，当无穷远时， O_1P 与 O_2P 变得平行，视差变成 0。



提示：让计算机自动建立多幅图像之间的匹配关系（如找到 p_1 点的对应点 p_2 ）其实是三维重建最困难的一个问题。图像匹配最常见的做法是基于特征点（如 Harris、DoG 特征点）的匹配，此外还有基于直线的匹配，以及基于区域的匹配。

基于立体视觉的直接三维重建

经典的三维重建方法是根据拍摄的左右两张图片，基于立体视觉直接进行形状重建（Shape from Stereo），一般都利用标定块事先计算出了摄像机的内参数。然后，找出 2D 图像上的对应点，并通过对对应点所求解得到的基本矩阵来计算摄像机的**本质矩阵（Essential Matrix）**，对于摄像机的运动参数恢复起最本质的作用，**其只与外参数有关**，据此可进一步分解得到摄像机的外参数和投影矩阵。有了投影矩阵就万事大吉了，如图 6-24 所示，根据三角测距原理，再采用 DLT（Direct Linear Transformation，直接线性变换）方法即可恢复空间点的 3D 高精度坐标。

基于分层的 SFM 三维重建

然而，很多时候我们拍的多张照片中并没有事先准备标定块，而且我们也并不要求特别高精度的 3D 重建，这时该怎么办？仔细观察图 6-22 和图 6-25 可发现，针孔相机（中心射影）拍摄的 2D 图像向 3D 欧氏空间的重建过程，实际上依次经历了从射影、仿射到相似变换的提升过程。实际上，相似变换是仿射变换的特例（子群），而仿射变换又是射影变换的特例（子群）。于是，**分层重建（Stratified Reconstruction）**作为一种代表性的 3D 重建策略^[17]，把原本多个变量的同时求解分摊到各个层中，这样减轻了非线性优化的难度，还可同时实现摄像机的自标定。针对连续拍摄的多张图像序列，分层重建首先利用图像对应完成**射影重建（Projective Reconstruction）**，再通过确定无穷远平面来完成**仿射重建（Affine Reconstruction）**，最后再通过

摄像机定标实现**相似重建 (Similarity Reconstruction)** 也即**度量重建 (Metric Reconstruction)**。步骤如下：

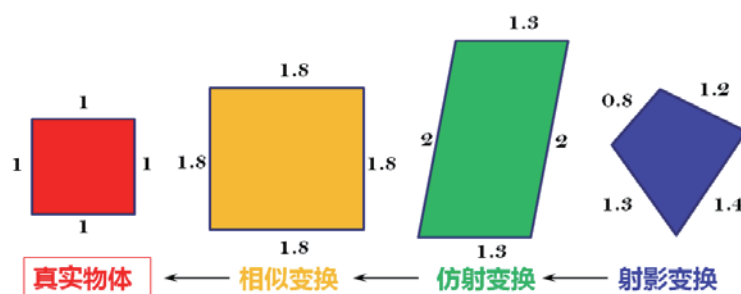


图 6-25 相似变换是仿射变换的特例，而仿射变换又是射影变换的特例

- 首先，仅根据图像的特征对应点我们就可得到射影重建，方法很简单。比如在双视图中，利用对应点求解（见公式（6））出的基本矩阵就完全描述了射影结构关系。注意射影意义下重建的点并不是实际物理点的度量重建，而是相差一个射影变换。
- 仿射重建是分层重建算法中至关重要的环节。可通过模约束（其至少需要 3 个视图）确定无穷远平面（如求解出它的法向量），之后就可计算无穷远平面的平面单应（Homography，可理解为射影变换）矩阵，将重建结果从射影重建提升（Upgrade）到仿射重建。
- 通过求解 IAC（参见 6.3.1 节的自标定）来计算出摄像机内参数，这样便可获得度量重建，其与真实大小物体的**欧氏重建 (Euclidean Reconstruction)** 只相差一个相似变换（均匀大小缩放）。

当然，分层重建并不意味着每层都必须经历，我们还可以选择跳层。比如利用**（对偶）绝对二次曲面**的成像（DIAC，是绝对二次曲线的像 IAC 的对偶），就可将三维重建直接从射影空间提升到度量空间。为了求解的稳定，可对摄像机内参数做一些合乎事实的约束限制，比如光心位于图像中心、畸变参数为零、纵横比 d_x/d_y 为 1 等。

我们往往拿着相机围绕固定物体转 360° 拍各个视角的多张照片，这时需要同时求解运动过程中的摄像机结构参数（包括不断变化的内外参数）以及固定物体的 3D 坐标，这被称为 **SFM (Structure from Motion, 基于运动的重建)**。由于双视图相机模型的参数求解存在一定的不确定性，所以 SFM 可采用更稳定可靠的多视图模型（比如用 3 张视图作为一组重建）。此外在求解时，为减少 3D 空间点重投影到图像平面的评估误差，需要把摄像机参数和 3D 空间点的位置都作为变量绑定在一起进行反复迭代优化，称之为**捆绑调整 (Bundle Adjustment)**，正所谓“东北乱炖一锅香”。

由于特征点匹配算法只能得到少数**稀疏 (Sparse)** 特征点的匹配对应，这样只能重建出物体的大致 3D 轮廓。所以还需要在此基础上对点（或面片）进行扩散，并根据极几何约束、视差梯度约束、颜色灰度一致性、错点剔除，以获得**稠密 (Dense)** 匹配重建或**准稠密 (Quasi-Dense)** 匹配重建。目前多视角立体重建的代表方法是先用 Bundler 软件包得到摄像机的内参数，然后再用 PMVS (Patch based Multi-View Stereopsis) 软件包获得稠密的 3D 点云。

当然，对于 3D 照相馆的室内影棚环境，我们完全可以事先对多个同步拍摄的摄像机用标定块进行标定，得到摄像机的参数，这样大大降低整个重建系统的难度。

基于 Visual Hull 的三维重建

上面介绍的基于立体视觉的三维重建方法，首先需要在多视图中找出特征点匹配，然后根据摄像机的参数来得到空间点的 3D 位置，最后才将 3D 点云拼接对齐并重建成 3D 网格面片。除此之外，还有一些方法直接采用多边形网格逼近物体表面，比如 **Visual Hull（可视外壳）** 方法。Visual Hull 是一种基于侧影轮廓的重建技术(Structure from Silhouette),原理非常简单。如图 6-26 所示，每个视点连接各自视图下的物体侧影轮廓，形成了一个视线锥体。然后，所有视点的轮廓锥体进行相交操作,所共同围成的空间包络就是可视外壳(图中的红色部分)。对于一个物体来说，其三维模型必定落在可视外壳中。注意，图中给出的只是一个 2D 示例，仅对 3D 物体的某个 2D 平面切层进行了重建，即假设图中的那条龙只是个平放着的 2D 图案。



图 6-26 上：Visual Hull（可视外壳）的重建原理；下：两个重建实例（图片来源：UIUC）

Visual Hull 简单快速，但也有明显的缺点，这就是精度不高。首先在原理上，有限个锥体相交得到的空间包络本身就不是原始形状的精确拟合，即会存在冗余体。当然，如果你拍摄更多视角的照片可明显改善。其次，对于那些非穿透的内凹部分也无法精确重建出来，比如深陷的眼窝就很难从轮廓中精细勾画出来。为了改进 Visual Hull 的重建效果，我们可以结合前面介绍的立体匹配重建方法，利用极几何约束、可见性分析、灰度梯度一致性、特征匹配等约束来获得一些精确的 3D 特征点，并以此驱动初始的 Visual Hull 不断收敛逼近到物体表面。

6.4 众里寻她千百度——海量3D模型的检索

现在网上的3D模型的数量可谓海量,仅 Shapeways 这一个网站就有 100 多万个模型。因此,绝大多数情况下我们根本无须亲手设计模型,只需把自己需要的找出来再 3D 打印即可。

“找”这个字说起来容易,做起来难。不像文本这样的结构化数据,3D 模型属于典型的非结构化数据。因此文本时代的网页搜索引擎已经力不从心了。有的读者会说,我给每一个 3D 模型用文本描述/标注一下不就行了?如图 6-27 所示,输入关键词 car,则将 3D 模型库中的所有的小汽车都找了出来。但如果是 100 万个模型,手工标注的工作量就太大了,比如这个模型要标上“一头脸上有道刀疤的黑色的猪”,那个模型要标上“一只黄色的短尾巴狗”等,十分烦琐。而且,光凭文本是无法描述清楚一个具体形状的。比如你钟情的一位美丽姑娘,你即使是用“娇洁凝碧水,似飘雪落轻盈;修弱纤纤指,柔声语嚶;巧笑倩兮,美目盼兮;夫何瑰逸之令姿,独旷世以秀群;佩鸣玉以比洁,齐幽兰以争芬”这样的一堆形容词来描述,别人也还是无法想象出她的具体样子来。

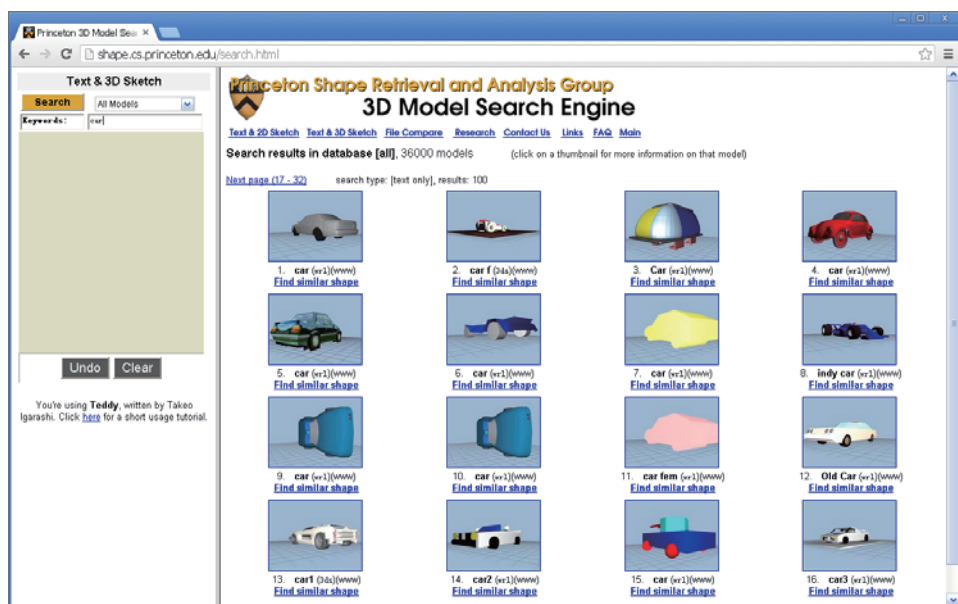


图 6-27 通过文本进行 3D 模型检索。左：输入关键词 car，右：返回所有的小汽车模型

因此,最靠谱的形状检索方法还是直接面向形状本身,让智能算法自动提取出形状的特征,然后计算形状的相似度度量,将最相似的结果返回给你。要把整个算法流程对之前没有任何基础的读者描述清楚,是个极具挑战性的艰巨任务。因此,下面笔者会把问题做不影响本质的简化,并一步一步娓娓道来。

6.4.1 线性分类与感知机模型

为阐述方便,我们把问题简化成两类。假设整个数据库共有 10 000 个模型,但只有两种类型的 3D 模型:人体的模型和猩猩的模型。在检索之前,我们先上传一个已有的模型(人体或猩猩),

接下来算法的目标是：要对此模型进行分析，判断它到底是人体还是猩猩。如果是人体，则将数据库中所有的人体模型列出来；反之，则把所有的猩猩模型列出来。于是，问题的关键在于：视觉算法如何智能地判别我们上传的那个 3D 模型到底是人体还是猩猩呢？

为了让视觉算法解答这个问题，第一步需要提取出形状的特征。我们这里讨论最直观的几何测量特征，比如，我们对数据库中所有模型都测量它们的身高、手臂长度、头围、两眼之间的距离、大腿长度、嘴巴突起的高度、门牙长度等 50 个特征。

考虑到数据库中有 $K=10\,000$ 个模型，50 个特征将形成很高维的数据量，即所谓的“**维数灾难**”（又名**维度之咒**，Curse of Dimensionality）。因此我们需要降维。最常用的方法是使用主成分分析 PCA，将不重要的或冗余的特征删除。为直观起见，我们假设重新生成的特征中只选择了最重要的两个：“手臂长度与身高的比例” θ 、“嘴巴突起的高度与头围的比例” δ 。



提示：采用主成分分析（PCA）降维后，得到的新特征未必有直观的几何解释，往往都是抽象的、不具备直观意义的。本例采用直观特征主要是为了解说方便。

维数灾难有两个角度的解释。当样本数量很大时，高维会导致数据量很大，占用大量的计算机内存和 CPU 处理时间。当样本数量较少时，高维又会使使得可用数据在空间中变得很稀疏，从很多角度看都不相似，导致分类器的预测能力降低，参数得不到正确的估计。

这样，数据库中的 10 000 个模型，每个模型都有两个特征。如果把每个模型都远看作一个点 \mathbf{x} （也即一个二维的特征向量），投影到二维屏幕上，每个点的 $x^{(1)}$ 和 $x^{(2)}$ 坐标对应那两个特征值（ θ 和 δ ），我们就得到了类似图 6-28 的一张图。其中，人体模型用红色的“+”号表示，代表正样本；猩猩模型用绿色的“-”号表示，代表负样本。

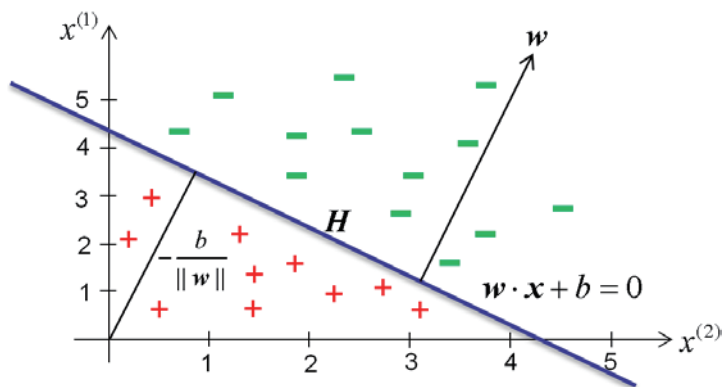


图 6-28 模型库中所有模型在二维平面的投影

我们在数据集中间画一条直线（图中蓝色的线），如果能够直接将这 10 000 个点划分成两类，那么这个数据集也被称为**线性可分的**（Linearly Separable）。如果把这条直线看作一个垂直于纸张的平面 H ，则其被称为**分离超平面**（Separating Hyperplane），简称**超平面**。

下面我们回到我们最开始的问题。用户提供一个 3D 模型，经过特征提取和降维后，在二维坐标上投影为一个二维点 \mathbf{x} （是个包含 $x^{(1)}$ 和 $x^{(2)}$ 两个坐标值的向量）。下面要判断点 \mathbf{x} 位

于超平面的哪一边，以此来确定到底是人体模型还是猩猩模型。

我们将上面的文本抽象为数学语言。具体来说，我们可以抽象为如下的函数，这个函数有一个很酷的名字：**感知机 (Perceptron)**，是**神经网络 (Neural Network)**与**支持向量机**的基础。从上面的描述可以知道，感知机是一个线性分类器 (Linear Classifier)。

$$y(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (7)$$



提示：是不是觉得数学符号看得很头疼？其实，抽象的数学所描绘的事物本质往往都是非常简单的，只不过符号化之后让人觉得不太直观和习惯而已。

其中， $\mathbf{w} \cdot \mathbf{x} + b = 0$ 就是超平面 H 的数学表达。 \mathbf{w} 是垂直于超平面的法线向量， b 是超平面的截距。超平面将特征空间划分为两类，在本例中，法向量指向的一侧为负类，另一侧为正类。

$\text{sign}(z)$ 为**符号函数**，只有两个返回值 (+1 和 -1)，在本例中分别代表是人体模型还是猩猩模型。

$$\text{sign}(z) = \begin{cases} +1, & z \geq 0 \\ -1, & z < 0 \end{cases}$$

有了感知机，我们就可以很方便地判别一个模型属于哪一个类别了。只需把一个新坐标值 \mathbf{x} 代入公式 (7) 中进行计算，如果位于超平面的左边， $\mathbf{w} \cdot \mathbf{x} + b \geq 0$ ，则结果 $y(\mathbf{x})$ 应为 +1，则知道这个模型是人体；如果位于超平面的右边， $\mathbf{w} \cdot \mathbf{x} + b < 0$ ，则结果 $y(\mathbf{x})$ 应为 -1，则知道这个模型为猩猩。OK，是不是特别简单？！



扩展：Novikoff 收敛定理。如果训练样本集是线性可分的，那么感知机算法可在有限次迭代后收敛。令 $R = \max \|\mathbf{x}_i\|$ ($1 \leq i \leq K$) 为输入样本特征向量的最大长度，超平面关于训练样本集的几何间隔距离为 γ ，则感知机算法在训练样本集上的误分类次数不超过 $(R/\gamma)^2$ ，也即最多做 $(R/\gamma)^2$ 次更新就会收敛。

证明：感知机算法仅在样本点分类错误时进行更新，令 \mathbf{w}_k 是第 k 次误分类时所更新的（法向量）权值，算法从 $\mathbf{w}_1 = \mathbf{0}$ 开始。假设第 k 次误分类发生在样本 (\mathbf{x}_i, y_i) 上，类别标签 $y_i = +1$ 或 -1 ，并将截距 b 归一化为 0，则有：

$$y_i(\mathbf{w}_k \cdot \mathbf{x}_i) \leq 0 \quad (8)$$

下面给出感知机的更新规则：当学习率 $\eta = 1$ 时，我们有 $\mathbf{w}_{k+1} = \mathbf{w}_k + y_i \mathbf{x}_i$ 。将两边都乘以单位长度法向量 \mathbf{w}_u ($\|\mathbf{w}_u\| = 1$)：

$$(\mathbf{w}_{k+1})^T \mathbf{w}_u = (\mathbf{w}_k)^T \mathbf{w}_u + y_i (\mathbf{x}_i)^T \mathbf{w}_u \geq (\mathbf{w}_k)^T \mathbf{w}_u + \gamma$$

因为初始 $\mathbf{w}_1 = \mathbf{0}$ ，将上式递推可得：

$$(\mathbf{w}_{k+1})^T \mathbf{w}_u \geq k\gamma \quad (9)$$

根据上面的公式 (8) 以及 $R = \max \|\mathbf{x}_i\|$ ，可得到：

$$\begin{aligned}\|\mathbf{w}_k\|^2 &= \|\mathbf{w}_k + y_i \mathbf{x}_i\|^2 = \|\mathbf{w}_k\|^2 + \|\mathbf{x}_i\|^2 + 2y_i(\mathbf{w}_k \cdot \mathbf{x}_i) \\ &\leq \|\mathbf{w}_k\|^2 + \|\mathbf{x}_i\|^2 \leq \|\mathbf{w}_k\|^2 + R^2\end{aligned}$$

同样，经递推可得到：

$$\|\mathbf{w}_{k+1}\|^2 \leq kR^2$$

两边开根号，并根据公式（9），可得：

$$\sqrt{k}R \leq \|\mathbf{w}_{k+1}\| \leq (\mathbf{w}_{k+1})^T \mathbf{w}_u \leq k\gamma$$

因此， $k \leq (R/\gamma)^2$ ，证毕。

通过上面这个定理可知，误分类的数目不会超过样本特征向量的最大长度与几何间隔的比值的平方。也即，当训练数据集线性可分时，感知机算法必然收敛（当不可分时，算法不能保证收敛，迭代结果会发生振荡）。此外，权值 \mathbf{w}_k 的整个更新过程实际上就是与新输入样本 \mathbf{x}_i 的线性组合，因此这是个**在线学习（Online Learning）**的过程，即根据新来的样例，边学习边给出结果。

6.4.2 支持向量机 SVM 与逻辑回归 LR

同时，我们从图 6-28 中可以看出，感知机的超平面不是唯一的，因为有无数个超平面都能把两类数据分开！那么，是否存在一个唯一的、最优的分离超平面呢？

答案是肯定的。我们能够找到这么一个超平面，它使得分类**间隔最大（Margin Maximization）**，这时的这个分类器称为**支持向量机（SVM, Support Vector Machine）**。

如图 6-29 所示， \mathbf{H} 就是这个最优的超平面，而 \mathbf{H}_1 、 \mathbf{H}_2 是两个与 \mathbf{H} 等距平行且通过最邻近的正 / 负样本点的平面。此时， \mathbf{H}_1 与 \mathbf{H}_2 之间的**间隔（Margin）**最大，其与最优超平面的法向量 \mathbf{w} 有关，间隔大小等于 $2/\|\mathbf{w}\|$ 。

我们把 \mathbf{H}_1 和 \mathbf{H}_2 通过的最邻近正 / 负样本点称为**支持向量（Support Vector）**。可以看出，在决定最优超平面时，**只有最外围的支持向量起作用，而其他内部样本点不起任何作用**（去掉这些点不会有任何影响）。因此，支持向量机是典型的“少数派报告”，只由很少的几个支持向量所决定。

讨论一下，支持向量机为何青睐最大间隔，却不是最小间隔呢？因为最大间隔能获得最大的稳定性与区分度。超平面之间的距离越大，分类器的推广能力越好，也就是预测精度越高。

有些聪明的读者可能会进一步问：“您刚才讨论的都是线性可分的情况，如果数据本身不是线性可分的，那超平面怎么去分开它们呢？”

Good question！实际上，现实情况很多时候就是线性不可分的，也即：此时的分类问题是非线性的。这里有两种解决方案。第一种解决方案针对**近似线性可分**的情况，把上面那样的**硬间隔（Hard Margin）**变成**软间隔（Soft Margin）**，也即引入**松弛变量（Slack Variable）**，允许超平面两边有**误分类点**，但应该让误分类点的个数尽可能小，同时让软间隔尽可能地大。

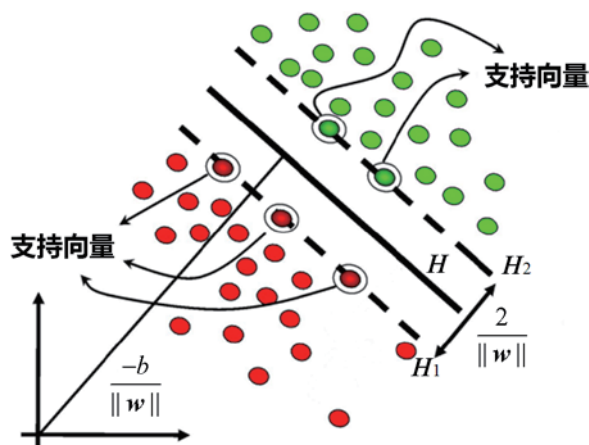


图 6-29 支持向量机与支持向量 (图片来源: Bülent Üstün)

另一种解决方案是针对**线性不可分**的训练数据, 我们可以利用核函数, 即通过非线性特征映射将输入变量映射到一个高维**特征空间** (它有一个很酷的名字叫希尔伯特空间: Hilbert Space), 在这个高维特征空间中构造最优分类超平面。“映射”这个动词可理解为变换或对应。如图 6-30 左边所示是一个分类问题, 我们无法用直线 (线性模型) 将正 / 负样本分开, 但可用椭圆 (非线性模型) 将它们分开。如图 6-30 右边所示, 如果我们采用某个**核函数 (Kernel Function)**, 通过非线性变换将原始样本映射到高维特征空间中去 (比如由 X/Y 二维映射到 $X/Y/Z$ 三维), 这时就可以使得原本的非线性问题简化为一个线性问题了 (通过一个超平面将样本分开)!

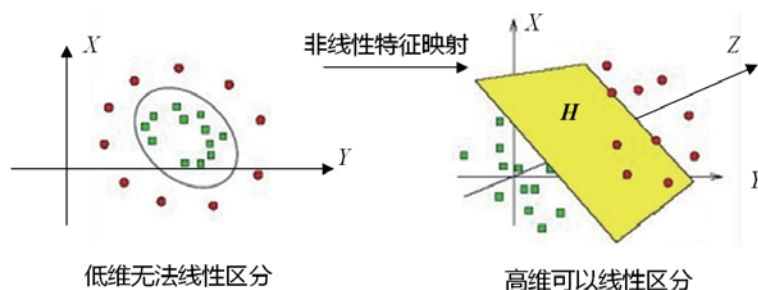


图 6-30 将低维线性不可分的数据, 映射到高维变成线性可分 (图片来源: DTREG)

提示: 再强调一下, 我们并不直接定义映射函数和 (高维) 特征空间的具体形式, 只需定义核函数即可, 这就是所谓的核技巧 (Kernel Trick)。



具体地, 设 \mathcal{X} 是输入空间 (比如本例中的二维空间 \mathbf{R}^2), \mathbf{H} 为 Hilbert 特征空间 (为高维空间, 如本例中的三维空间 \mathbf{R}^3), 如果存在一个从 \mathcal{X} 到 \mathbf{H} 的映射 $\phi(x): \mathcal{X} \rightarrow \mathbf{H}$, 使得对任意 $a, b \in \mathcal{X}$, 函数 $K(a, b) = \phi(a) \cdot \phi(b)$ 。这里的 $K(a, b)$ 就是核函数, $\phi()$ 为映射函数, $\phi(a) \cdot \phi(b)$ 中间的圆点符号 \cdot 代表**内积**运算。

上面之所以考虑内积运算, 是因为如果把支持向量机转化到对偶形式 (见第 10 章 10.1.3 节), 可发现**分类函数 (超平面 H)** 的计算只涉及输入样本之间的**内积运算**, 甚至不必知道映射函数 $\phi()$ 的具体形式。实际上, 通过定义映射函数 $\phi()$ 去获得高维的

内积通常并非易事，而通过定义核函数 $K(a, b)$ 在映射到高维同时可方便地获得内积——对于任意的对称函数 $K(a, b)$ ，只要满足 Mercer 条件，就可作为内积使用。支持向量机采用恰当的内积函数（核函数）便可实现经某一非线性映射后的线性分类，而计算复杂度却没有增加，因为这个高维空间中的线性分类器与空间的维数无关。

然而，如何针对不同的问题选择不同的核函数仍是一个悬而未决的问题，往往依赖于领域知识。下面是常见的一些核函数。

- 多项式核函数（Polynomial Kernel Function）： $K(a, b) = (a \cdot b + 1)^p$ ($p > 0$)，所对应的支持向量机是一个 p 次多项式分类器。
- S 形（Sigmoid）核函数： $K(a, b) = \tanh(\beta(a \cdot b) + \gamma)$ ，其中双曲正切函数 \tanh 是 Sigmoid 函数的一种，所对应的支持向量机是两层感知机的一种特例。
- 高斯核函数（Gaussian Kernel Function）： $K(a, b) = \exp(-\frac{\|a - b\|^2}{2\sigma^2})$ ，所对应的支持向量机是高斯径向基函数（RBF, Radial Basis Function）分类器。

上面这段文字仍比较拗口，怎么理解它呢？我们可以将左图理解为散布在面团上的红、绿两种颜色的豆子，在二维平面上无法一刀切开（只能通过画一个非线性的椭圆）。这时，我们可以揪住内部的面团往下拉（即非线性映射），这样绿色方形豆子都被移动到红色圆形豆子的 Z 轴下方了，由于所处高度不同，此时红、绿两种颜色的豆子就可以一刀切开了。



提示：非线性映射是针对整个样本空间的，本例中指整个面团。这里我们定义的核函数，使得内部的面团受到的向下拉力更大，能够比外围的面团更快地下落到 Z 轴下方。

还有更聪明的读者会进一步问道：“您刚才讨论的都是两类问题（人体、猩猩），如果是多类问题（人体、猩猩、大象、老虎、狮子）呢，怎么办？”

Very good question！其实，多类问题可通过转化为多个两类问题来求解。比如共有 5 个类。我们可以第 1 次问“是第 1 类还是第 2 类”，第 2 次问“是第 1 类还是第 3 类”，第 3 次“是第 1 类还是第 4 类”，如此问下去，你终能知道属于哪一个类别。但如此一来，我们总共必须设计“ $5 \times 4 / 2 = 10$ ”个分类器。即在 M 个类别的情况下，总的两类分类器数目为 $M(M-1)/2$ 。



扩展：除了 SVM，在工业界实际用得较多的一种机器学习方法是逻辑回归（Logistic Regression, LR），用于估计某种事物的可能性（取值 0 ~ 1 之间）。与 SVM 不同，逻辑回归既可用于分类（按照可能性大于 0.5 或小于 0.5 来区分），也可用于回归。逻辑回归仅能用于线性问题，其本质实际上仅在线性回归的基础上，套用了一个 Sigmoid（S 形）函数，如图 6-31 所示。

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

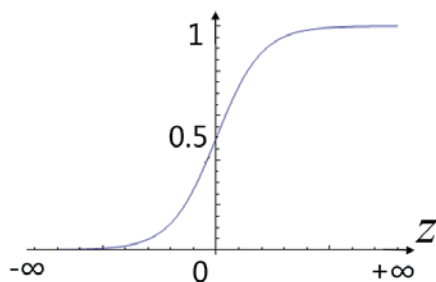


图 6-31 Sigmoid (S 形) 函数

Sigmoid 函数 $\sigma(z)$ 的优势在于，虽然自变量范围是从 $-\infty$ 到 $+\infty$ ，但**值域范围限制在 0~1 之间**，也即可被看成一个概率函数。这种归一化被认为是合乎“逻辑”的，并在实际场合中得到广泛应用。

另一方面， z 是多个特征变量的加权线性组合：

$$z = w_0x_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n。$$

比如一个女生对你的好感程度（打分）可用逻辑回归量化到一个 0~1 之间的数值，其由多个线性因素加权得到，如通过计算你的年龄大小 x_0 、经济收入多少 x_1 、……、身材高度 x_n 等各方面条件综合得到；而 w_0 、 w_1 、…、 w_n 是她分别为年龄、经济收入、身高等特征设置的权值参数（回归系数），分别代表着在她心目中对每个特征的看重程度。因此，为了预测她对你的打分，关键就是要通过她之前为多名特征各异男士的已有打分，来估计出她心目中对各项特征条件的权值系数 w_0 、 w_1 、…、 w_n 。在具体实现时，可将问题形式化为最大似然估计（参见第 10 章 10.8.3 节），并采用梯度下降（参见第 10 章 10.3.1 节）优化方法进行求解。综上，**逻辑回归实质为一个线性分类模型，它与线性回归的不同点在于：逻辑回归将原本线性回归输出的很大范围的数，例如从负无穷到正无穷，限制在了 0 ~ 1 之间。**

6.4.3 基于内容的 3D 模型检索

前面我们已经提到过，我们希望进行基于内容的 3D 模型检索（Content-based 3D Retrieval），也即不是根据文本描述，而是根据模型本身的几何形状、拓扑结构等进行自动检索。



提示：拓扑学（Topology）是几何学的一个分支，不涉及长度和角度的测量，研究的是当形状发生连续形变时，那些不会改变的性质，这种变换称之为拓扑变换或**同胚（Homeomorphism）**。比如，我们对形状进行伸缩和弯曲（但不能撕破和黏合）后，拓扑是不变的，即同胚的。这个过程如同手捏有弹性的橡胶膜一样，因此拓扑几何也被称为“橡胶膜几何”。拓扑学不区分咖啡杯和甜甜圈，因为一个足够柔软的甜甜圈可以捏成咖啡杯的形状，两者都是有 1 个孔的、即**亏格（Genus）**为 1 的形状。同理，正方形和圆（都有 1 条**边界 Boundary**）、立方体和球面（都有 0 条边界），也分别都是同胚的，它们都是亏格为 0 的形状，即它们的**欧拉示性数（Euler Characteristic）** $2-2g=2-2\times 0=2$ 。

在用户界面上，基于内容的 3D 模型检索有两种交互模式。

- 3D 示例模型检索。现有的大多数检索系统都只提供 3D 示例模型检索：即要求用户上传一个示例模型。比如你上传一头猪的 3D 模型，则检索系统将模型库中所有猪的 3D 模型作为返回结果。这种交互模式还可扩展到 2D 示例图片检索，即用户只需提供一张拍摄有一头猪的 2D 照片，系统也能返回模型库中所有猪的 3D 模型。
- 手画草图检索。手绘 2D 草图（如图 6-32 所示）和手绘 3D 草图（如图 6-33 所示）的检索方式提供了比较友好的界面，用户可以通过手绘 2D 或 3D 草图来表达检索要求。然而缺点是对用户的手绘技巧有一定要求，不能画得太差。

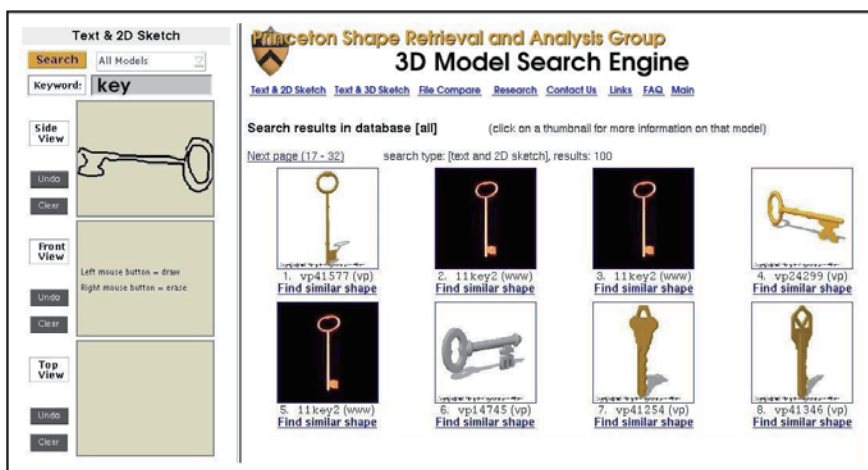


图 6-32 2D 草图检索（通过在左边手绘一个 2D 的钥匙来检索模型库中的 3D 钥匙模型）
（图片来源：普林斯顿大学）

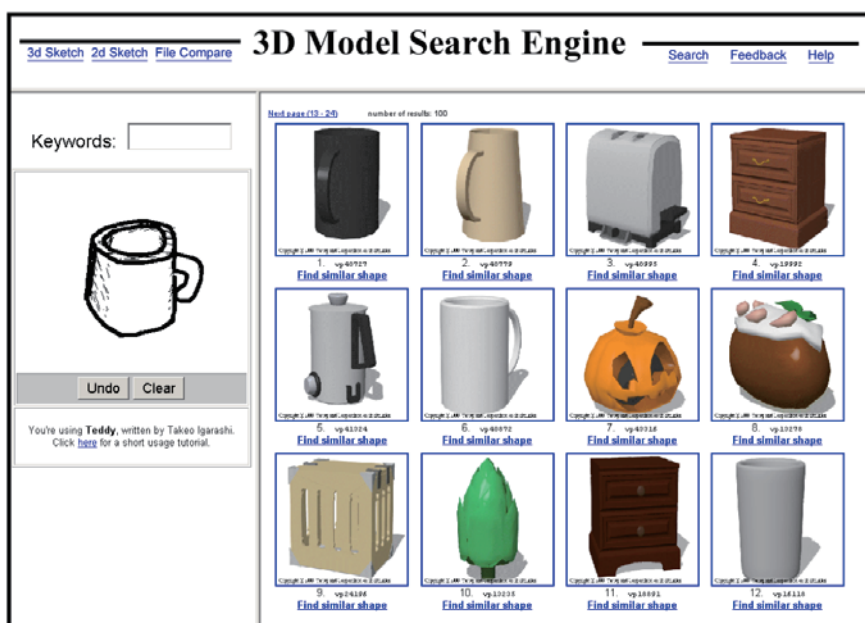


图 6-33 3D 草图检索（通过在左边手绘一个 3D 的茶杯来检索模型库中的 3D 茶杯模型）

为了让检索系统能更清楚地了解用户的意图，还可引入相关反馈，即通过人机交互，让用户不停地“告诉”计算机：检索结果中哪些是符合要求的（相关模型）、哪些是不符合要求的（不相关模型）。系统根据用户的反馈来对用户的个人偏好进行分析，对检索结果进行修正，并返回新的结果给用户评价。如此反复交互，直到用户满意为止。

在技术原理上，基于内容的3D模型检索主要有两个步骤：提取3D模型的特征，然后比较所提取特征的相似度，以便将那些与示例模型相似度大的模型作为结果返回。

一个理想的**形状特征描述符（Shape Descriptor）**应具有几何不变性（即对模型的平移、旋转、缩放等具有不变性）以及拓扑不变性（比如你双手抱拳与双手分开时的身体拓扑是不同的，但你仍是你）。此外，形状描述符对于噪声影响、模型简化/细分、姿态变化等也应该是保持不变的。如果按照所提取的3D模型特征的不同，3D模型检索技术可以划分为这么几大类。

- 基于形状的检索技术。大多数3D模型检索技术都是基于形状特征的，可进一步分为基于空间域的检索和基于频率域的检索。基于空间域的检索技术大多数使用了统计学方法，即通过统计数据来描述模型的形状特征，如形状直方图、形状分布（Shape Distributions）等。基于统计的特征对于模型的区分度不够好，因而只适用于模型的粗分类。基于频率域的检索技术可借助傅里叶变换、球面调和、流形调和（Manifold Harmonics）等频谱数学工具。
- 基于拓扑结构的检索技术。如采用骨架或中轴线这种拓扑结构来描述有关节的模型（如人体），代表性的有 Multi-Dimensional Reeb Graphs、Shock Graphs 等。拓扑结构特征比较符合人眼对模型的全局直观感受，但缺点是特征提取和匹配的时间开销过大。
- 基于二维投影图像比较的检索技术。将3D模型往某个侧面投影，得到一幅投影图像，这样就可直接参考已经非常成熟的2D图像检索方法了。代表性的方法有 Spin Images、Shape Contexts 等。注意：二维投影的图像会受到旋转及缩放的影响，因此在投影之前需要对模型进行归一化处理。

接下来的**相似性度量**就是计算（多维）特征向量之间的相似度，即计算用户上传的示例模型的特征向量 f 与三维模型库中某个模型的特征向量 f_i 之间的相似性距离。常见的距离度量有 Euclidean 距离、Manhattan 距离、Hausdorff 距离、Earth Mover's distance（推土机距离）、Pyramid Match Kernels（层次匹配核）等。距离越小，说明两个模型的相似性程度越高；反之，距离越大，说明两个模型之间的相似程度越小。如图 6-34 所示，最终系统会根据相似性度量的距离大小进行排序以返回查询结果，从而实现基于内容的模型检索。检索好坏的常见评价指标有：查准率（Precision）和查全率（Recall），即生成所谓的 PR（Precision-Recall）曲线。













Match # : 0 Distance : 0.000 	Match # : 1 Distance : 0.028 	Match # : 2 Distance : 0.185 	Match # : 3 Distance : 0.418 	Match # : 4 Distance : 0.626 	Match # : 5 Distance : 0.682 
Match # : 6 Distance : 0.750 	Match # : 7 Distance : 0.769 	Match # : 8 Distance : 0.802 	Match # : 9 Distance : 0.813 	Match # : 10 Distance : 0.856 	Match # : 11 Distance : 0.875 

图 6-34 根据相似性度量的距离大小进行排序返回查询结果

此外,按照检索对象的不同,现有的三维模型检索方法还可分为“全局检索”和“**局部检索**”^[30]两大类。如图 6-35 所示,如果用户仅将一把椅子的靠背(属于椅子的局部)作为示例模型(操作上只需用鼠标框选住靠背部分),则检索系统将返回局部检索结果,即返回数据库中所有跟这个靠背(而不是整把椅子)相似的模型。

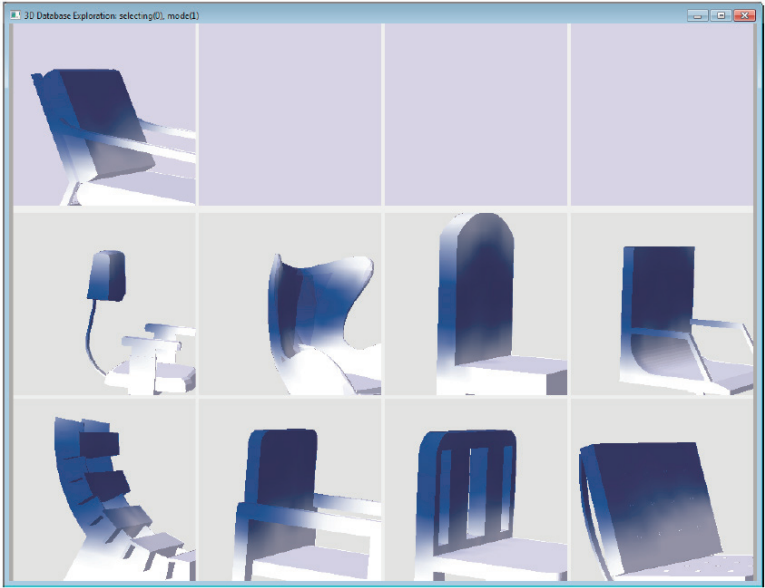


图 6-35 局部检索的例子

目前网上已经有一些 3D 模型搜索引擎,如 Thingiverse、3D Partsource、Defcad.com、Yeggi、Yobi3d、Fabforall、Dimensionext、Google 3D 模型库等,你可以去它们的网站上体验一下。

6.5 形状拆解：大尺寸物件的自动分块打印

3D 打印的成本正变得越来越便宜，操作也正变得越来越简单。它已被广泛应用，但仍有使用局限：打印对象必须受限于 3D 打印机的打印空间。例如，家用小型 3D 打印机 Cube 虽然售价仅为 1 299 美元，但它只能打印一个午餐盒大小的东西。MakerBot 2 能打印的最大尺寸也仅为：28cm×15cm×15cm。目前家用 3D 打印机难道就只能满足于打印玩具士兵、橡胶印章和其他小型塑料饰品这类小玩意吗？那岂不是太没劲了？实际上，大型工业级打印机能打印的最大尺寸也大不了很多，比如曾被评为世界最大的 3D 打印机的 Objet 1000 也就只能打印 100cm×80cm×50cm，如图 6-36 所示。



图 6-36 曾被评为世界最大的 3D 打印机的 Objet 1000（图片来源：Stratasys）

看来，目前 3D 打印机本身确实难以提高打印尺寸，除非你愿意花费很高的价格让厂商专门订做一台。聪明的读者会说：我想到了一个好办法，把 3D 模型分割成几个小块分别打印，然后再组装起来不就 OK 了？这确实是个好办法！但是，如果 3D 模型比较大，要手工拆分成大量的零部件，就会变得很烦琐。而且你能保证这些分块组装在一起就结实吗？比如你打印组装了一把大椅子，你真的敢每天坐在上面而不担心垮塌吗？

怎么办？如果你还记得 3D 打印有一个孪生兄弟，一切就会迎刃而解。是的，让 3D 智能数字化来解决这个问题！不用花大价钱去买大型的工业级打印机，也不用担心强度不够。

普林斯顿大学的研究团队研发了一款智能软件 Chopper^[32]，它能够将一个尺寸大于 3D 打印机的模型，就像它的名字那样“剁（Chop）”成若干块，同时，软件会自动生成连接节点（类似于插销和凹槽），方便用户组装和黏合分散的部件，以便最终拼装成完整物品，如图 6-37 所示。



图 6-37 Chopper 软件自动将尺寸大于 3D 打印机的模型分成多个部件（图片来源：普林斯顿大学）

算法在设计连接节点时，尽可能使连接节点远离受力点，且接缝隐蔽，结构坚固，并以最少分块次数为佳。要优化这样的算法并不容易，但是 Chopper 现在基本上能够计算出比手工设计更优化的方案。

下面展示了几个案例：Objet Connex 500 打印机打印的母子雕像（注意观察各部件之间的连接节点，如图 6-38 左边所示），Fortus 400mc 打印机打印的猫咪（如图 6-39 所示）。



图 6-38 Objet Connex 500 打印的母子雕像



图 6-39 Fortus 400mc 打印的猫咪

6.6 形状分析：优化桌面3D打印机打印精度的表现力

首先我们来观看著名的**克拉尼金属板（Chladni plates）**实验，如图 6-40 所示。18 世纪，德国物理学家及音乐家克拉尼在薄的金属板上撒上均匀的细沙，然后拉起小提琴。在声音的振动下，这些细沙开始“跳舞”，并从一些地方向另一些地方聚集，并最终形成各种对称的图案。克拉尼发现不同频率的声波会形成不同的图案，且频率越高则形成的图案越复杂！

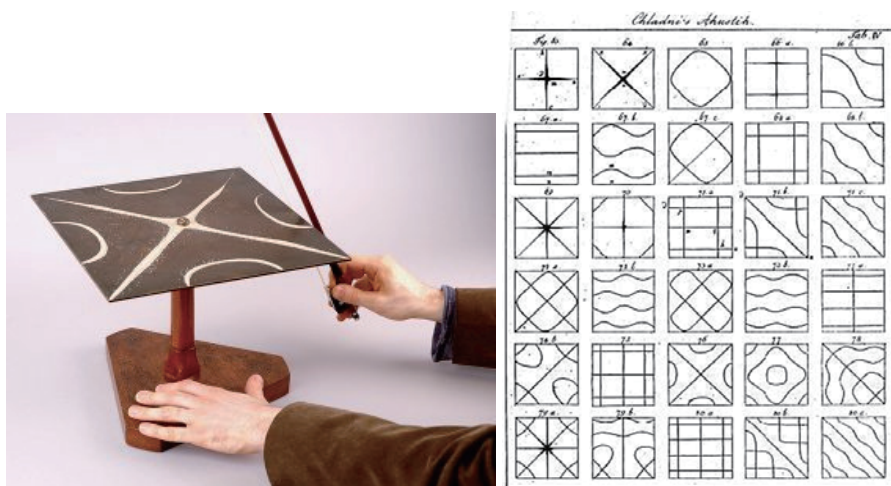


图 6-40 克拉尼金属板（Chladni plates）实验（图片来源：The Whipple Museum）

由此我们知道：声音原来是“有形状的”！其实很早以前，古埃及人就把几何叫作“冻结的音乐”，他们用几何学来记录乐谱。甚至有人大胆地猜测：通过听鼓发出的声音频率，能猜出鼓的形状。（“Can one hear the Shape of a Drum?”）。虽然答案是“并不能完全对应”（存在着不唯一性），但至少可以看出：形状和频率有着极其紧密的联系。

非常类似于二维平板上的克拉尼图案，在 3D 形状上不同频率也对应于不同的条纹分布，如图 6-41 所示。同样，频率越高，则条纹的分布情况也越复杂（从左到右）。此外，我们还可发现，这些频谱条纹像是“理解了”整个形状似的，不仅随着整体形状的走势有规律地自然延展、一点儿也不错错综杂乱，而且在形状对称的地方也呈现出对称的形态（见左腿和右腿区域的条纹）。我们将定义在任意**流形**形状的这种频谱分布称之为**流形调和（Manifold Harmonics）**^[33]。

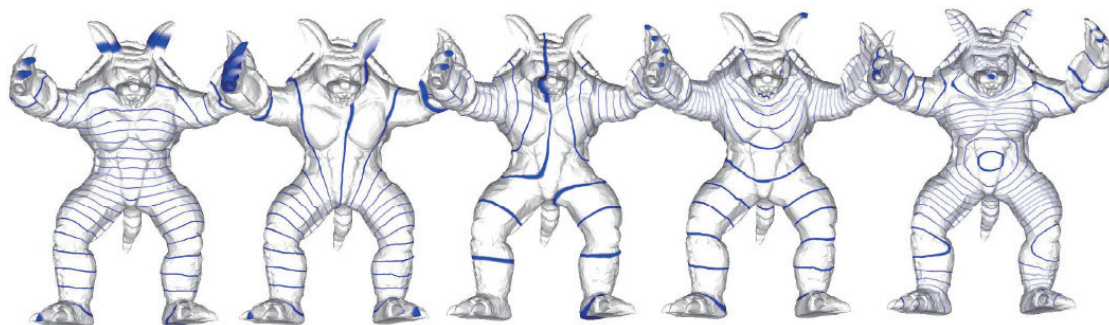
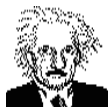


图 6-41 3D 形状上的频谱条纹（图片来源：Bruno Lévy）



提示：流形调和是定义在任意流形曲面上的 Laplace-Beltrami 微分算子的**特征函数 (Eigenfunction)**，将其二维平面上的离散傅里叶变换和球面上的球调和函数扩展到了三维离散曲面。

我们知道，函数的 Laplace (拉普拉斯) 算子定义为它的梯度的散度，即

$$\Delta = \text{div grad} = \nabla \cdot \nabla = \sum_i \frac{\partial^2}{(\partial x_i)^2}。$$

Laplace-Beltrami 微分算子则是规则域上 Laplace 算子在流形曲面（非规则域）上的推广，即：

$$\Delta = \text{div grad} = \frac{1}{\sqrt{|G|}} \sum_i \frac{\partial}{\partial x_i} (\sqrt{|G|} \sum_j g^{ij} \frac{\partial}{\partial x_j})。$$

Laplace-Beltrami 微分算子的特征模态（包括特征函数和特征值）满足 Helmholtz 波动方程： $\Delta f = -\lambda f$ ，其中标量 λ 称为 Δ 所对应的**特征值 (Eigenvalue)**，在有的文献里也被称为**形状 DNA**，解 f 称为对应于 λ 的特征函数。上式经过离散化后，可得**特征值和特征函数对** (λ_k, H^k) ，其满足 $-\Delta H^k = \lambda_k H^k$ 。通过**流形调和变换 (Manifold Harmonic Transformation, MHT)**，原本的**空(间)域**坐标就可被转换成**频(谱)域**坐标，即 $\tilde{x}_k = \langle x, H^k \rangle$ 。

如果我们进一步研究，还可以发现一个更有趣的现象。如图 6-42 所示，我们把图中的条纹图变成更精细的彩色图(由于排版原因,这里只取前3个频率分量)。本图中实际上有两个 3D 模型，其中下方的模型是对上方的 3D 模型做了一个细节保持的形变（参见本章 6.2.3 节），也就是局部形状细节是刚性保持的、**等长不变的 (Isometry Invariant)**。请仔细观察，可以发现：形变前后的这两个模型的频谱分布是非常相似的、几乎可以说是等同的！这个有用的性质可用于 3D 模型的检索（参见本章 6.4.3 节），以确保将形变前后的形状归类为同一个形状。

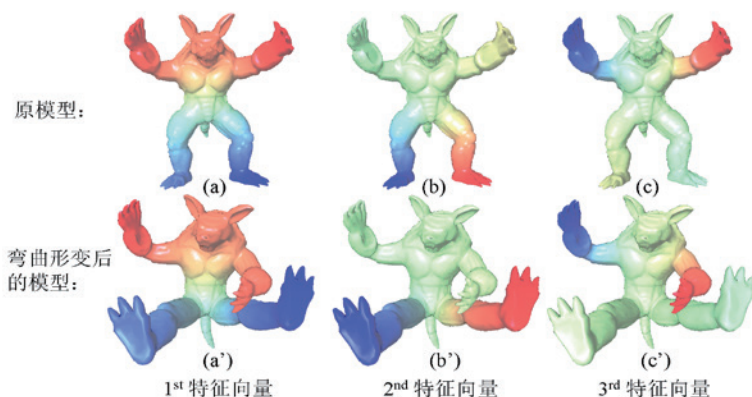


图 6-42 形变前后的形状频谱分布保持相似

再举一个 3D 形状的例子，如图 6-43 所示。图 (a) 显示了原始的 3D 形状，图 (b) 显示了只保留形状低频部分的光滑效果（使用低通滤波函数），图 (c) 显示了只保留高频部分的细节增强效果（使用高通滤波函数），图 (d) 显示了对某些频带进行特别处理后的夸张效果（使用带通滤波函数）。

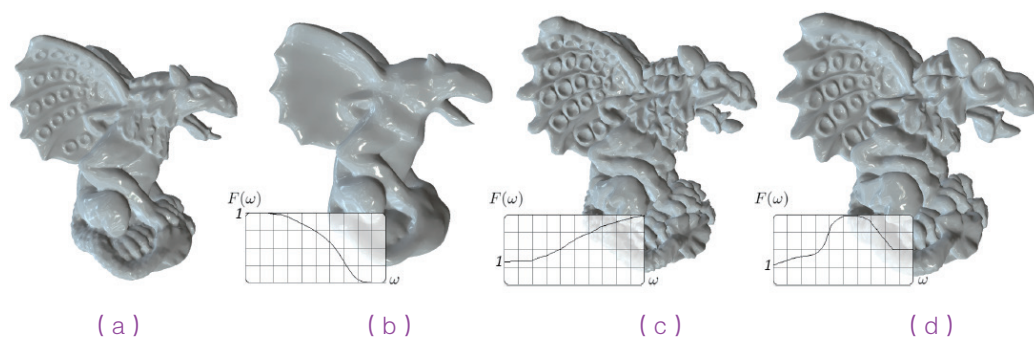


图 6-43 3D 形状的滤波

因此，利用频谱我们可以对形状的表现力做一个层次化的展示。如图 6-44 所示，从左到右分别对应一个 3D 头像的低频和高频部分。可以看出，形状的低频部分看起来非常平滑，但缺乏细节；而形状的高频部分则细节丰富、立体感强。

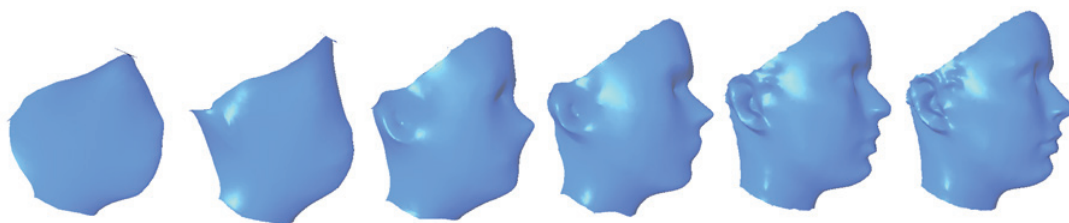


图 6-44 基于频谱的形状细节层次化展示

我们知道，当前 3D 打印的主要矛盾在于有限的打印设备精度和用户期待的理想打印结果之间存在着较大的差距。而通过对 3D 数字形状进行智能分析将有效地缓解这一矛盾。比如通过频谱分析，可对 3D 形状的频域特征空间进行智能化处理，优化生成最匹配于当前打印机精度的 3D 数字化模型（如选用图中的某个中间状态的 3D 头像模型）。

6.7 形状平衡：如何确保3D物件站立稳当

在虚拟的 3D 数字环境下建模，你创作的模型可以不平衡，甚至可以摆出违反重力原则的造型，这都不是问题。但是，当把模型打印成实物，如果形状不平衡，你可能就需要把它粘在很重的基座上了，甚至对它进行修改，以便使模型能够立起来。这个令人头疼的问题现在被一个叫“站起来”（Make it Stand^[34]）的智能软件很巧妙地解决了。“站起来”通过快速简单地优化模型来使模型保持平衡——这样即使你随意放置，也无须额外的支架或底座，模型也不会倒！如图 6-45 所示，“站起来”软件使原本需要两条腿和一条尾巴共同支撑才能站立的马儿得以“金鸡独立”。



图 6-45 从左到右：“站起来”使原本需要两条腿和尾巴一起支撑才能站立的马得以单腿站立
(图片来源：ETH Zurich)

下面进行具体解释。用户首先根据重力原则把 3D 模型摆出一个大体的造型，设置接触点。接触点可以是模型和地面接触的地方，也可以是用绳子吊起来时的接触点。所有的接触点构成了一个接触平面，称作支撑多边形。要让模型保持平衡的关键在于：模型的重心的垂直投影必须落在支撑多边形内！



提示：支撑多边形的严格定义是模型与地面的所有接触点的凸包 (Convex Hull)。凸包是计算几何 (Computational Geometry) 中的一个概念，其定义为：点集 Q 的凸包是指一个最小凸多边形，使得 Q 中的点要么在多边形边上要么在其内。凸包的一个有趣且优良的性质是：凸包内部的任意一点都可表示为凸包顶点（如图 6-46 左边所示最外围的那 6 个顶点）的凸组合（见第 10 章 10.1.1 节），即可通过凸包顶点的线性组合来得到。如图 6-46 左边所示是二维的例子，蓝色的凸多边形就是凸包。在本节中我们指的就是二维凸包。当然，我们也可以把凸包推广到三维，如图 6-46 右边所示的红色三角形曲面。

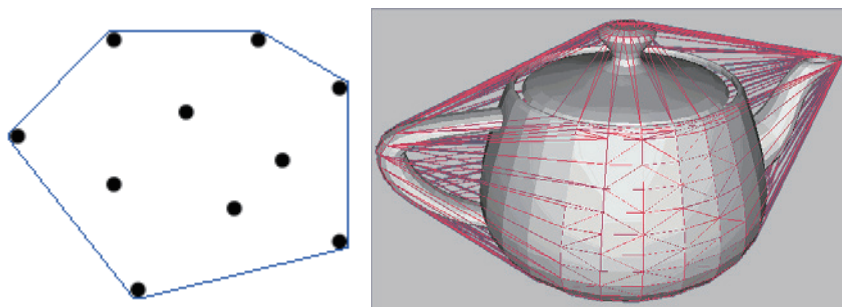


图 6-46 二维和三维的凸包 (Convex Hull) 示例 (图片来源：Autodesk)

OK，掌握了这个原则，下面的一切就交给智能化算法好了。具体来说：软件通过使用很多个细小的体素来构建 3D 模型内部，一个体素要么让它空着，要么把它打印填充满。为了保持平衡，

软件选择性地挖空体素（如图 6-47 所示的黄色部分），以便把重心的投影稳定地落在支撑多边形内。当然，由于挖空了部分体素，这样同时也节省了打印耗材。

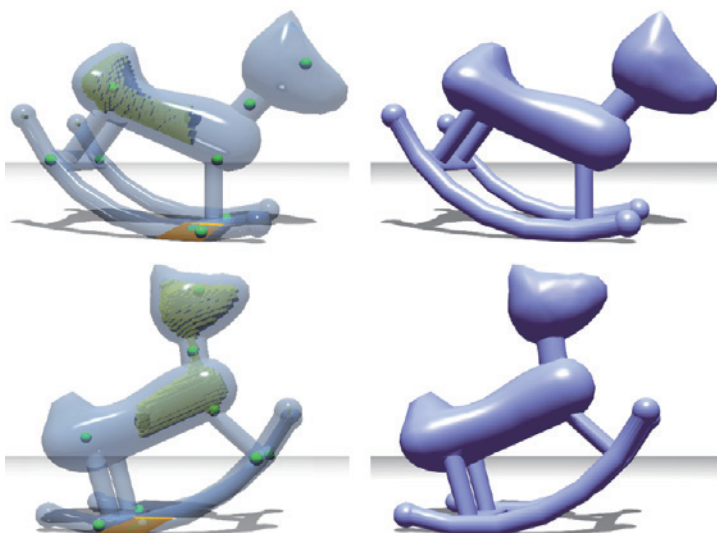


图 6-47 通过调整玩具摇摇马的内部填充改变它的平衡（上图是前腿平衡着地，下图是后腿平衡着地）



提示：一般在计算机中显示的 3D 模型都是三角形网格“曲面”，而不是“体”，也即内部是空的。但要用 3D 打印机打印实体模型，就需要考虑内部的填充和支撑了。

然而，通过挖空部分体素的办法并不是万能的，这时我们可以采用另一种思路：稍微改变一下模型的造型。千万别小看这种肉眼几乎看不出来的改变，你可以回忆一下大学军训时站军姿：看上去几乎一样的站姿，你隔段时间把身体稍稍往前或后倾一点点，就会感觉轻松稳当很多。改变模型的造型需要用到形变编辑技术，如前面 6.2.3 节所述，形变的关键约束在于细节保持（Detail Preserving），这样才能保证模型细节不失真，使得形变后看上去仍是原模型的样子。通过在模型上交互地操纵几个手柄节点，我们可以对 3D 模型的特定部分进行平移、缩放和旋转。如图 6-48 所示，我们对原始模型做了细微的编辑，几乎观察不到变化，但却使得模型可以很好地保持站立。这种方法和上一种体素挖空方法协同作用，效果更好。



图 6-48 从左到右：原始模型（重心的投影落在支撑多边形外）、编辑后的模型（重心的投影落在支撑多边形内）、3D 打印结果。编辑后的模型几乎看不出变化，但却可稳稳地单脚站立

6.8 形状优化：生成坚固的内部轻质结构使得耗材最省

受原材料和工作原理的限制，当前的 3D 打印技术具有耗材昂贵、机械寿命短、加工速度慢等缺点。这些都不是一时半会就能快速解决的。不过，我们可以把目光转向 3D 模型本身，通过 3D 智能数字化技术来优化模型的形状，即可达到立竿见影的特效！这样，在减少所需耗材的同时，自然也加快了加工速度，同时也减少了机械损耗。

这里介绍一下中国科学技术大学研究小组提出的一种基于“蒙皮 - 刚架”（Skin-Frame）的轻质结构^[35]。如图 6-49 所示，这里的刚架（Frame）实际上类似于建筑中常见的桁架（Truss）；不过刚架的节点是固定的，而桁架的节点是活动的。刚架结构能有效地降低打印材料成本，并使打印物体满足所要求的物理强度、受力稳定性、自平衡性及可打印性。这些刚架是由一些细杆通过一些节点相连而成的，形成空间的一个图结构。这种结构的优点主要有两个：一、力学特性好，当某节点受到外力时，此处的受力能通过相邻的细杆迅速传播分散开来；二、质量轻便，这种结构是由稀疏的细杆组成，因此总体质量不大，很好地减少了结构本身的重量及所使用的材料。

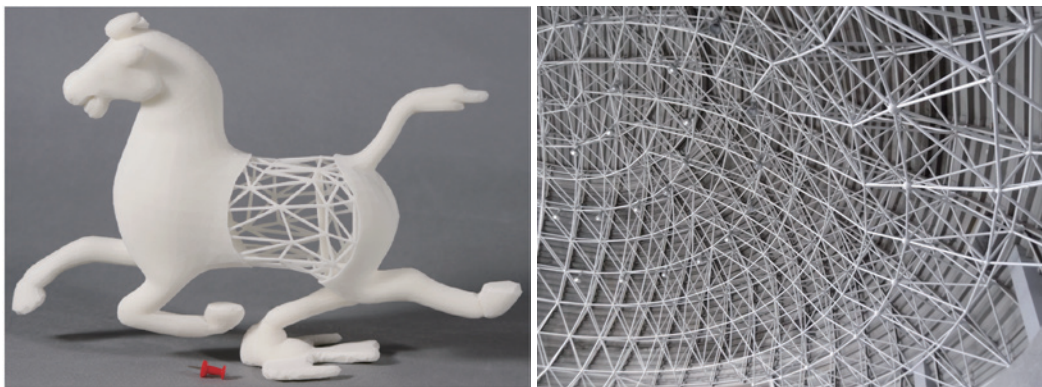


图 6-49 “蒙皮 - 刚架”（Skin-Frame）轻质结构（左）与建筑行业的桁架结构（右）类似
（图片来源：中国科学技术大学）

可能有读者会说，我也实际打印过一些物件，目前主流的 3D 打印机都采用直接掏空中间的办法来节省耗材，方法比这种简单而且也挺省耗材的。确实，使模型中空确实可以省材料，但到底挖空多少（也即外表皮到底剩多薄）却是个很不好掌控的量。挖得多了，外表皮太薄，则表面强度不够；挖得少了，跟实心差不了多少，还是不太节省材料。还有的读者会说，我的打印机还可以在内腔中打印蜂窝状的结构，类似于 6.7 节中提到的体素，看着也很结实。是的，蜂窝结构确实是大自然的杰作，但密密麻麻地布满了空腔，所以综合算下来也不是最省材料的（当然，我们也可在受力结实的地方多挖空一些，在受力薄弱的地方，比如细细的脖子部位，少挖空一些或者干脆不挖空，这种策略称之为自适应受力的空心化）。而在本节中介绍的“蒙皮 - 刚架”结构，如图 6-50 所示，外边只需薄薄的一层表皮，里面的刚架经过拓扑和几何优化之后，变得非常稀疏，远没有蜂窝那样密密匝匝，这样极大地节省了耗材，同时刚架结构又充分保证了强度。

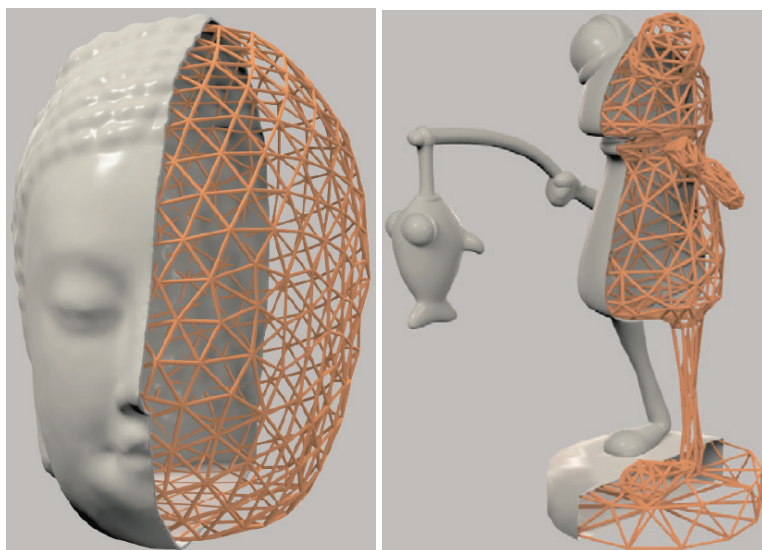


图 6-50 放大显示的“蒙皮 - 刚架”轻质结构（图片来源：中国科学技术大学）

为了获得最优的轻质结构,智能算法的目标函数包括两个。第一个目标要使物体的体积最小,即蒙皮体积及刚架结构的体积之和最小。由于蒙皮的厚度的增加会很快增加体积,因此将蒙皮的厚度固定为最小可打印精度,不作为优化变量。因此,需要优化的变量只包括刚架结构中的细杆的半径、节点的个数及位置。第二个目标为使得刚架结构中的细杆数量及节点数量尽量少,该目标是为了不要出现冗余的细杆及节点。为此,可进行多目标优化建模,并采用迭代优化的数学方法来优化这两个目标函数。

智能数字化算法的另一个优势是,还可以让计算机优化每一根刚架的粗细,甚至可以让每一根刚架的粗细都各不相同!而这在现实中的“大规模量产”化的建筑行业显然是有心无力的,所以这也再次证明了 3D 智能数字化打印可实现“大规模定制”的优势,哪怕是在一个 3D 模型内部刚架的细节上。举个例子,如图 6-51 所示,右图的底座也比左图的底座短很多,如果每根刚架都一样粗细的话,则右图的模型会站立不住而倒下。如果我们把右图底座部分的刚架弄粗一些(越红表示越粗,越蓝表示越细),则可以让重心偏移,这样模型就可以保持平衡了。

除了在物体内部设计刚架结构外,该工作还考虑了外部的支撑结构的优化设计。经过第 3 章的打印实践,相信你对打印悬垂或中空部分时采用的支撑并不陌生,这样做的目的是防止悬垂部分在打印过程中塌下来,如图 6-52 左边所示。在第 2 章中我们讲过,这是 FDM(熔融沉积成型)打印机的工艺所决定的,因为它不像粉末式打印机那样有粉床的自然支撑。在打印完成后,这些支撑都要去掉,变成废料。这些支撑材料其实并不容易剥除掉,有可能会破坏物体本身的表面。更重要的是,支撑材料不是很便宜,所以看着左图中密密麻麻的支撑你的心是不是在滴血?没关系,再次让刚架节省你的 Money!如图 6-52 右边所示,看着是不是小清新了很多?支撑材料也随之节省了很多。

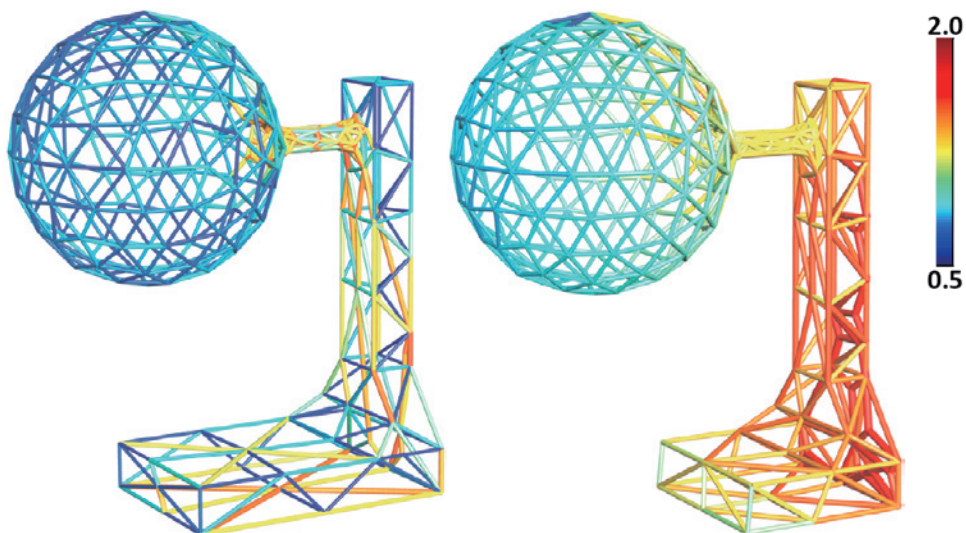


图 6-51 不同粗细的刚架（越红表示越粗，越蓝表示越细）（图片来源：中国科学技术大学）



图 6-52 刚架结构可大大节省支撑材料（图片来源：中国科学技术大学）

研究小组的实验结果表明：该方法对于某些物体比实心打印能节省约 70% 的材料，打印后的模型也轻了很多。如果你对此还没有直观的感觉，请再看一个打印案例。图 6-53 中的香蕉小人的双腿特别得细，为了不让它折断，以前的方法需要在香蕉小人身上固定一个支撑。但这样一根“拐杖”显然破坏了香蕉小人的美感和自尊。而采用“蒙皮 - 刚架”，可极大地节省了材料，模型也变得很轻，因此无须任何支撑也完全可以让香蕉小人自由地站立起来。

为了使 3D 打印出来的模型足够结实，我们还可进一步做形状力学分析。普渡大学和纽约大学的两个研究组分别进行了相关研究，如图 6-54 所示，他们不仅对重力负载、手捏施力（图 6-54 下图左一）以及最可能的破坏受力情况（图 6-54 上图中间）进行了力学分析，还通过模态分析（Modal Analysis）的方法并利用类似于模态叠加的原理找出那些最不结实的结构区域（Worst-Case Structure、Weakness Map），然后有针对性地对这些区域进行加强处理，如加粗增厚（图 6-54 下图左二）、掏空减重（图 6-54 下图右一）或附加支撑（Strut，图 6-54 下图右一）。

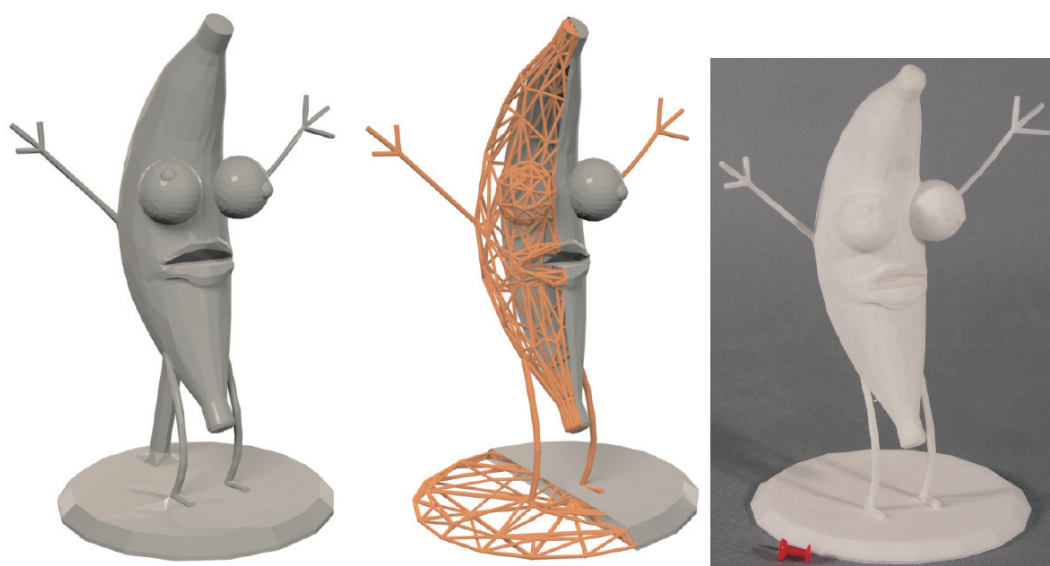


图 6-53 轻质结构大大减轻了模型重量。左图：以前方法需要添加支撑；中图：“蒙皮－刚架”结构重量轻、强度高，无须支撑；右图：3D 打印出来的模型（图片来源：中国科学技术大学）

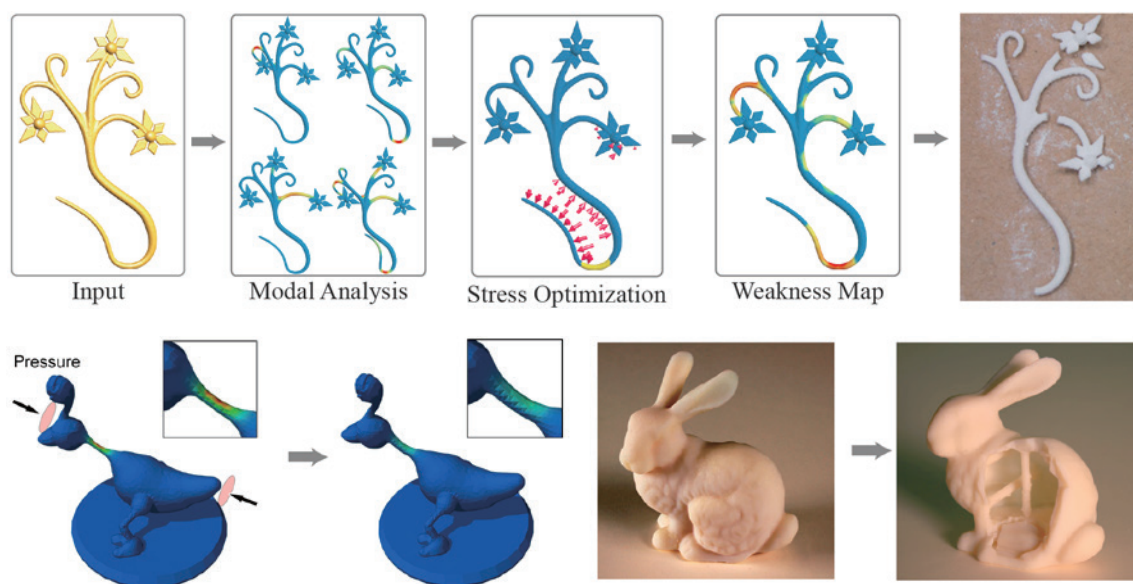


图 6-54 通过对重力负载、手捏施力（下图左一），以及最可能的破坏受力情况（上图中间）进行力学分析，找出那些最不结实的结构区域，并有针对性地对这些区域进行加强处理，如加粗增厚（下图左二）、掏空减轻（下图右一）或附加支撑（下图右一）（图片来源：纽约大学、普渡大学）

6.9 基于笔画的3D建模：让新手和孩子轻松设计形状

不可否认的是，对于普通用户而言，当前的 3D 数字化建模技术仍然门槛较高。主要原因是现在人跟计算机打交道一般都是通过 2D 屏幕上的鼠标光标。用只有 2 个自由度的 2D 光标去操

作高达 6 个自由度的 3D 模型，显然是勉为其难。于是我们不得不一次次地把 3D 模型旋转平移，以便让某一个操作面能够正对着我们。在一次次的旋转平移中，大多数普通用户终于被转得晕头转向，对 3D 建模心生畏惧。

有的读者会说：我已阅读过前面的章节，第 4 章 4.2.1 节的“‘所想即所得’：3D 设计的新境界”中不是介绍了一种 3D 鸟标吗？可以让用户在 3D 空间直接建模。没错！带有浓重虚拟现实色彩的 3D 鸟标确实大大拉近了人与数字空间的距离。但是，目前鸟标还是比较昂贵的，跟中关村里一个鼠标才卖 10 元相比，一时还难以走进千家万户。

那怎么解决这个难题呢？还是那句老话：让智能数字化来解决！通过视觉计算，用户只需像涂鸦一样，在 2D 屏幕上画一些 2D 草图，计算机即可根据这些 2D 笔画自动生成 3D 模型。下面分别介绍 3 款笔画式建模工具，以飨读者。

6.9.1 Doodle3D：3D 设计就像涂鸦一样简单

3D 打印机现在越来越流行，但是问题是，只有“技术宅”才知道怎样去摆弄它。Thingiverse 提供了数以千计的开源 3D 设计模型，给予你共享和使用。但是，它们依然不是你自己的设计！

传统纸笔模式的手绘草图是一种很自然的人机交互方式，手绘草图可以捕捉那些不完备但又稍纵即逝的思维灵感，防止过早地将模糊概念确定化（那样容易丢失模糊性中所蕴涵的无限可能性）。Doodle3D 就是这样一个类似于手绘草图的速写工具，用于创建极具个性化的 3D 模型。Doodle3D 非常容易使用，谁都能用它，直接用手指在触摸屏上随意画，然后软件自动帮你把 3D 模型轮廓补充出来，如图 6-55 所示。



图 6-55 Doodle3D 的操作界面和打印成果

如图 6-56 所示，你还可以像雕刻陶瓷一样雕刻你的作品，甚至可以轻轻地扭曲和旋转某一层。



图 6-56 Doodle3D 还可像雕刻陶瓷一样雕刻作品

6.9.2 Teddy/FiberMesh : 更精准的 3D 笔画建模

Doodle3D 虽然很容易上手,但实在太简单了,不能对 3D 模型做比较精细的编辑和修改。下面我就介绍一下 Teddy、FiberMesh 等系统(类似的还有 Shapeshop、Autodesk 123D Creature 等),可基于手绘草图进行比较精准的 3D 建模。操作步骤跟 Doodle3D 一样,用户随意地在 2D 屏幕上画几笔,然后系统根据这些笔画生成一个光滑的 3D 模型。主要原理是:通过从用户绘制的封闭笔画中抽取骨架,再根据轮廓点到骨架的距离进行膨胀生成三维网格曲面,并且膨胀的厚度可以交互控制,使得生成的模型具有多样性。此外,这些笔画还可作为控制曲线,以便对 3D 模型的轮廓和外形做进一步的调整,如剪除、追加、挖洞等,如图 6-57 所示。

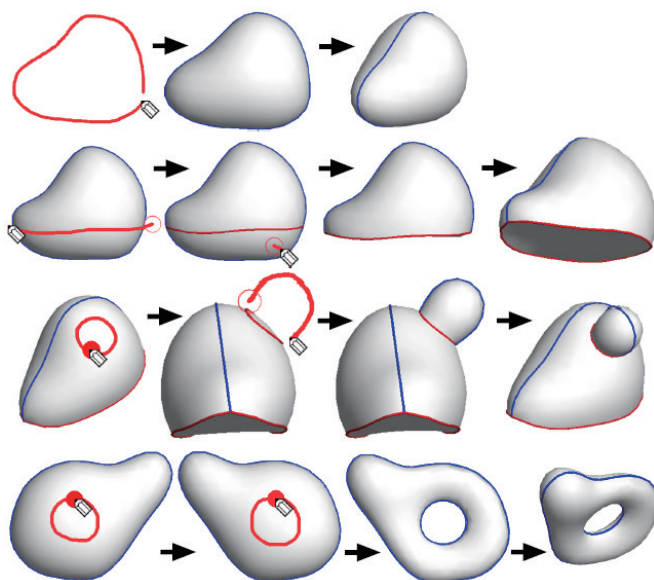


图 6-57 FiberMesh 的 4 种操作模式(创建、剪除、追加、挖洞)

这类建模方法尤其适合创建圆乎乎的 3D 玩具模型,如猫、狗、猪等卡通动物,如图 6-58 所示。非常适合用于儿童的即兴涂鸦,并将作品 3D 打印出来给孩子的家长。如果打印耗材选用的是巧克力,则还可以让家长一起来分享带有孩子劳动成果的美味。



图 6-58 Teddy 软件生成的卡通动物 3D 模型

6.9.3 3-Sweep 技术：轻松让照片中的 2D 物体变 3D 模型

除了通过手绘 2D 草图来进行 3D 建模,还有一种更加简单直观的方法,那就是你可以照着现成的二维照片直接描!清华大学的研究人员最近开发出一种名为“3-Sweep”的技术^[38],可以实现从单张 2D 照片直接生成 3D 模型,让 3D 建模变得像在 Photoshop 中建立选区、编辑图像一样简单。

图 6-59 中的图 (a) 给出了一张含有多个灯架的 2D 照片。视觉算法首先把所有灯架的 2D 边缘自动提取出来 (如图 (b) 所示)。然后,用户在某个灯架上画 3 条笔画 (图 (c) 中的红 / 绿 / 蓝线),对应于 X/Y/Z 轴 3 个空间方向,以完成物体的轮廓勾勒,计算机就可根据这些交互信息把这个灯架的 3D 模型恢复出来了。对所有灯架做同样操作,即可获得完整的 3D 模型 (如图 (d) 所示)。我们还可以对每个 3D 灯架做各种角度的任意旋转,并将编辑后的 3D 模型渲染到新的场景下,来生成一张全新角度的照片 (如图 (e) 所示)。

更详细的交互过程如图 6-60 所示。用户只需在屏幕上用鼠标画 3 条笔画 (红 / 绿 / 蓝),对应于 X/Y/Z 轴 3 个空间方向。中间的图是重建好的 3D 模型,最右边的图是将 3D 模型旋转到新视角再渲染成一张新的 2D 照片。

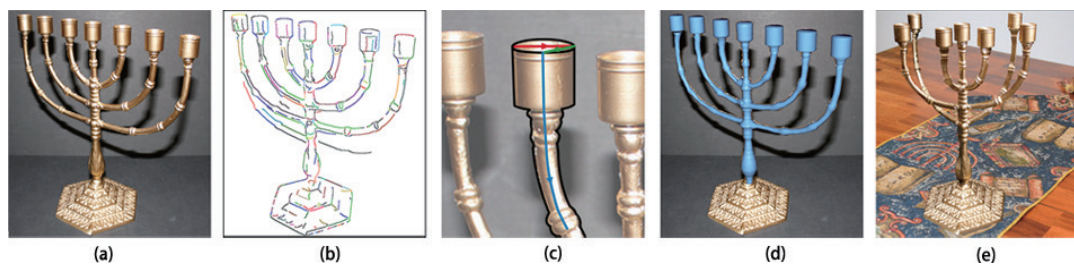


图 6-59 “3-Sweep” 的工作流程 (图片来源 : 清华大学)

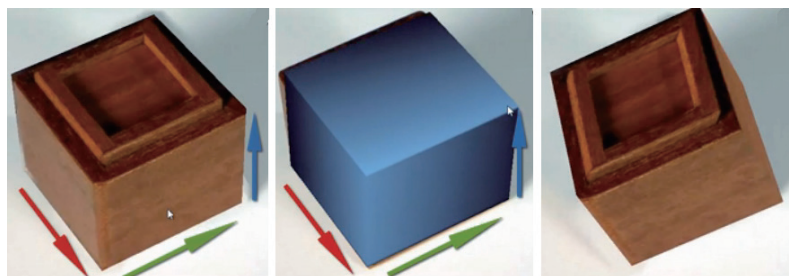


图 6-60 “3-Sweep” 的交互过程 (图片来源 : 清华大学)

“3-Sweep” 支持的建模物体需要能够被立方体、圆柱体或球体等这几种基本形状块所表示。由于 2D 转 3D 的不确定性, 单一分块的建模其实是不确定的, 这就需要依靠多个分块之间的垂直、平行等几何约束来确定三维模型。

除了可以生成模型外, “3-Sweep” 还可以直接为模型贴图, 并且在旋转编辑 3D 模型时, 还可以自动补全原本被模型遮挡住的背景缺失区域。这里采用的是基于 PatchMatch 的图像补全算法, 其类似于 Photoshop 的内容自动填充 (Content-Aware Fill) 功能。

更多的 3D 重建结果如图 6-61 所示, 效果是不是很震撼?



图 6-61 “3-Sweep” 的更多操作结果 (图片来源 : 清华大学)

由于算法假设了物体的对称性, 所以这项技术目前比较适合于轴对称的人造物体, 此外, 对模型的复杂性, 以及照片边缘的清晰程度也有一定的要求。

6.9.4 “神笔马良” 3Doodler：用笔直接画出 3D 线框实物

前面介绍的都是用笔画来进行 3D 数字化建模，下面介绍一种更直接的方法：直接用笔来画出 3D 实物！当然，靠人的手工是难以逐层去堆积每个面的，那样的话工作量太大。因此，这里只是用笔来画线，画出物体的线框图。

WobbleWorks 的团队在 Kickstarter 上发布了自己的产品 3Doodler —— 3D 涂鸦笔，用户可以拿着并自由地悬空作画，如图 6-62 所示。我们来看看 3Doodler 的一些基本的参数。这支笔的质量为 200g，可适用于 110V 或者 220V 电压环境。涂鸦笔的“颜料”则是 3mm 粗的 ABS 或者 PLA 塑料。而由于使用时会喷出高温热塑（270℃），所以并不推荐 12 岁以下的儿童使用，而用户使用的时候也不能触碰喷口，以免烧伤。喷出的热塑性材料会马上被内置的小风扇所冷却。

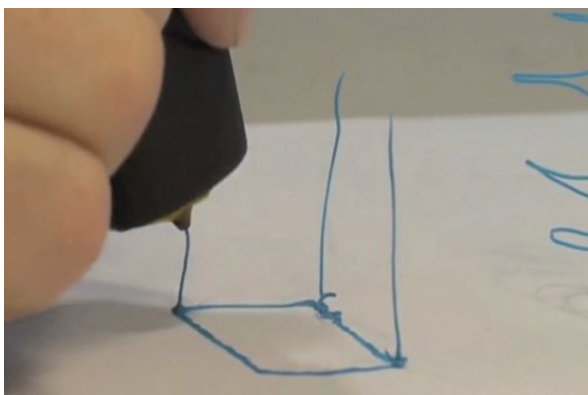


图 6-62 3Doodler 的操作过程

3Doodler 内部的构造其实非常简单，相当于从普通的 3D 打印机上取下了热挤压头，然后整合进一支笔中，ABS 塑料丝从笔的末端进入，在笔身中穿过并被融化，然后从笔头涌出，最后在较短的时间内完全凝固定型。

在操作时，用户可以事先设计平面草图（如图 6-63 左图所示），然后根据图纸涂鸦出每个部分，最后把它们组装起来——如巴黎埃菲尔铁塔（如图 6-63 右图所示）。

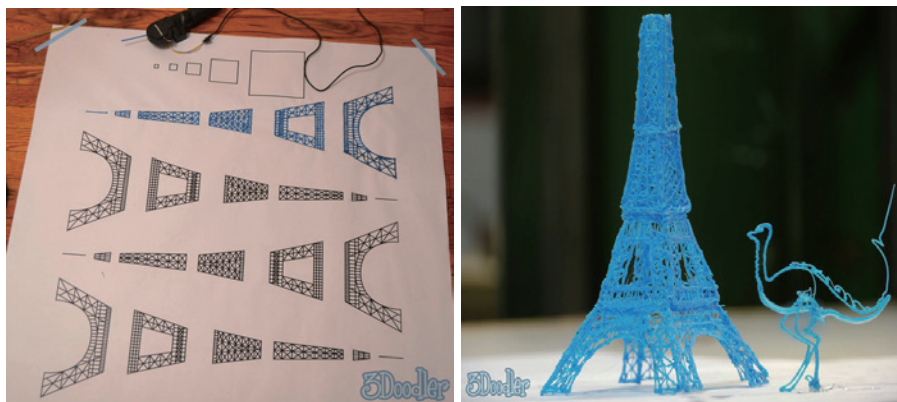


图 6-63 根据事先设计的平面草图进行 3D 创作

几乎每个人在自己的少年时期都拥有过一个梦想，那就是可以将自己在纸上画出的图形、物件变为现实。从过去很长一段时间以来，这一想法都仅仅停留在“梦想阶段”。3Doodler 让这一想法成为现实，而且不需要技巧，不需要专业技能，甚至不需要预先画好草图（当然有草图做出来的立体涂鸦更加规整美观）就可制作 3D 模型。

用户可以使用 3Doodler 制作各种 3D 形状模型：艺术品、首饰、装饰品，甚至是个性化的日常用品（例如，iPhone 或者手提电脑的保护壳），已经有一些艺术家使用这支“马良的神笔”创造出一些有趣的作品，例如，立体造型的动物和人，如图 6-64 所示。

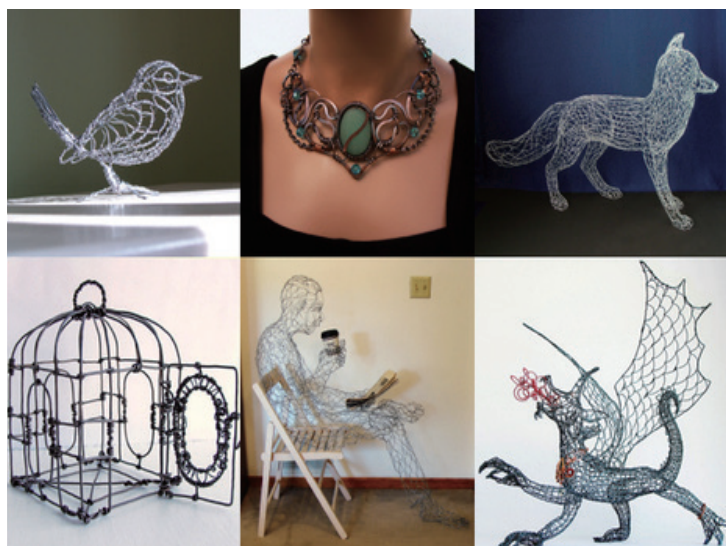


图 6-64 3Doodler 的作品示例

6.10 增强现实：在打印之前看到融入环境的真实效果

现在网上可免费下载的 3D 模型有很多，虽然在计算机里显示的样子很讨人喜欢，但真正打印成实物摆在房间里，到底跟环境搭不搭，就很难说了。这就跟去商场买一件衣服一样，放在衣服架子上很好看，但一穿在你身上，颜色和款式就未必适合你了。所以，如果在打印之前，能够预览一下 3D 模型与环境搭配的实际效果就完美了。

答案是肯定的，这要用到增强现实技术。**增强现实 (Augmented Reality, AR)**，是在**虚拟现实 (Virtual Reality, VR)** 的基础上发展起来的新技术，也被称为**混合现实**。它通过计算机技术，将计算机生成的虚拟物体、场景或系统提示信息叠加到真实场景中，从而实现对现实的增强。增强现实不仅展现了真实世界的信息，而且将虚拟的信息同时显示出来，两种信息相互补充、叠加。

这里介绍一款软件应用：Augment，有 Android 和 iOS 版，可让用户使用手机或平板电脑看到 3D 模型打印后放在手上的最终效果。比如预览一个沙发打印后摆在房间里的样子，如图 6-65 所示。



图 6-65 使用平板电脑预览一个沙发打印后摆在房间里的样子（图片来源：augmentedev）

当你把将要打印的 3D 模型载入 Augment，就可以先从各个角度旋转查看这个模型的样子，如图 6-66 所示。通过摄像头实时拍摄的真实背景，用户可以 360° 观察 3D 模型是否真正与环境搭配，非常实用。

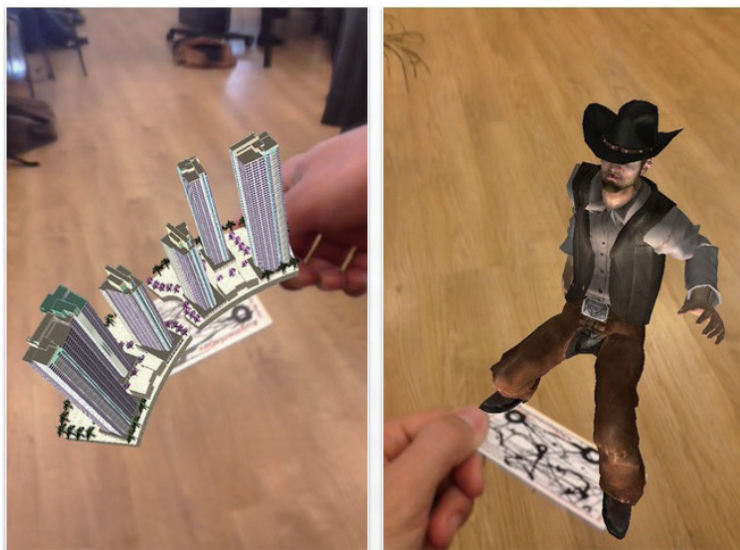


图 6-66 通过增强现实技术全方位查看 3D 模型

6.11 OpenCV与OpenGL：视觉计算入门的两大利器

通过本章的阅读，你可能已经对视觉计算产生了兴趣，并着手准备基于它来构建你的 3D 打印杀手级应用了。但应该怎么入门呢？还有更重要的是，如何尽快地动手实践，以便对视觉计算有一个直观的认识。这里给大家推荐 OpenCV 与 OpenGL，分别是计算机视觉和计算机图形学的实战利器，下面一一进行介绍。

6.11.1 OpenCV 与 AdaBoost 人脸检测

OpenCV 的全称是：Open Source Computer Vision Library，Intel 公司支持的开源计算机视觉库，采用 C/C++ 编写，可以运行在 Linux/Windows/Mac 等操作系统上。其目标是构建一个简单易用的计算机视觉框架，以帮助开发人员便捷地设计复杂的计算机视觉相关应用程序。OpenCV 包含的函数有 500 多个，覆盖了如工厂产品检测、医学成像、信息安全、用户界面、摄像机标定、立体视觉和机器人等领域。

下面我们举个例子，展示如何使用 OpenCV 自动检测图像中的人脸。正如前面所提到的，我们要实现个性化的 3D 打印定制，比如要为 1 万名用户量身定制眼镜，我们就需要利用 OpenCV 自动地将这 1 万名用户的人脸分别从照片中检测出来，进而对每个人的眼距、眼眶大小进行自动测量。当然，你还可以让智能算法自动分析每个人的脸型（圆脸、方脸、瓜子脸等等），因为不同的眼镜形状（如方形、圆形眼镜）适于不同的脸型。这样，每个用户的眼镜都不是一模一样的，这些量身定制出来的眼镜可以把每个用户的脸庞都装饰得美观、时尚。可以看出，智能算法是实现低成本 3D 打印“大规模定制”的前提和基础，因为要为 1 万名用户手工定制眼镜将是一项昂贵、且几乎是不可能完成的任务（假设工期仅限 1 周或 1 个月的话）。

看了上面这段文字，你是不是有些跃跃欲试？好，我们就正式开讲如何检测图像中的人脸。首先，我们介绍如何用 OpenCV 的寥寥数行代码即可实现读取并显示图像，代码如下。

程序 6-1：从文件中读取一幅图像并在屏幕上显示

```
#include "highgui.h"
int main(int argc, char** argv)
{
    if(argc<2)
        exit(1);
    // 读入一张图片
    IplImage* image = cvLoadImage(argv[1]);
    if (NULL == image) // 如果读入失败，退出程序
        exit(1);
    // 创建一个窗口，标题为 Example
    cvNamedWindow("Example", CV_WINDOW_AUTOSIZE);
    // 在窗口 Example 中显示图片 image
    cvShowImage("Example", image);
    // 暂停程序，等待用户触发一个按键
    cvWaitKey(0);
    // 释放图像所分配的内存
    cvReleaseImage(&image);
    // 销毁窗口
    cvDestroyWindow("Example");
    return 0;
}
```

图像在屏幕上的显示效果如图 6-67 所示。

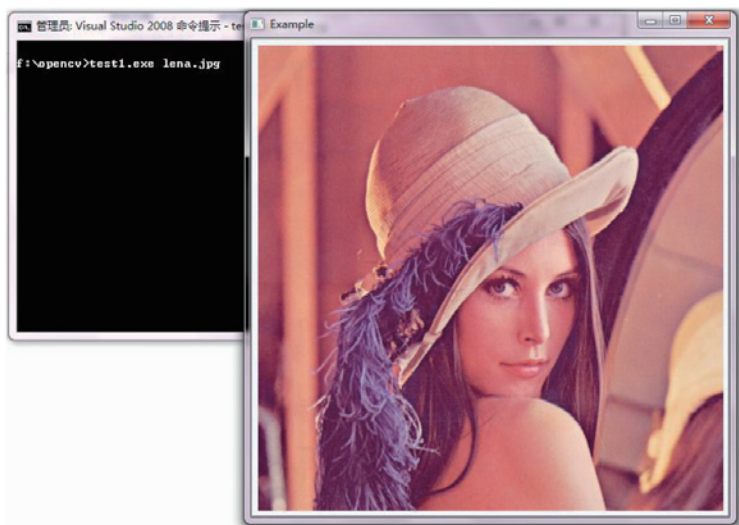
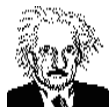


图 6-67 使用 OpenCV 读入一幅图像文件并在屏幕上显示

接着我们看看如何使用 OpenCV 来进行人脸检测，主要是调用 OpenCV 中已训练好的 Haar 分类器来对输入图像进行模式匹配。

人脸检测算法的流程非常简单，具体介绍如下：首先，构造一个含有成百上千张样本图像的样本库。网上现在有很多这样的图像库可供下载。接着，我们对每一张样本图像做标记，标记为正样本或负样本：正样本就是含有人脸的图像，负样本就是不含人脸的图像（如一张风景照片或一头猪的照片）。比如，我们这个样本库中含有正样本 700 张，负样本 300 张。然后，我们对每张图像提取 Haar-like 特征（请参考本章 6.2.1 节“个性特征的描述与检测”）。我们知道，Haar-like 特征是非常原始、粗粒度的。所以，需要对这些提取到的特征用 **AdaBoost（Adaptive Boosting，自适应增强、自适应提升）学习算法**进行特征选择和分类器训练^[23]。具体来说，首先从大量的特征中选出少量的关键特征，构成**弱分类器**（指一个分类器对一组数据的分类正确率只比随机瞎猜好一点，大于 50% 即可）。然后，通过类似跨国公司“兼并收购”小公司的方式，把这些弱分类器（按照各自正确率的大小）加权线性组合成**强分类器**（指一个分类器的正确率很高）。



提示：弱学习和强学习等价定理表明：只要有足够的数据，弱分类器就能通过不断组合的方式增强（Boost）为任意高精度的强分类器，正所谓“三个臭皮匠顶一个诸葛亮”！

在把弱分类器组合成强分类器的过程中，AdaBoost 方法的**自适应（Adaptive）**在于：前一个分类器分错的样本会被用来训练下一个分类器。AdaBoost 方法就是这样一种反复迭代的算法，在每一轮中加入一个新的弱分类器，直至达到某个预定的足够小的错误率。每一个训练样本都被赋予一个权重，表明它被某个分类器选入训练集的概率（即可能性）。如果某个样本已经被准确地分类，那么在构造下一个训练集中，它被选中的概率就被降低；相反，如果某个样本没有被准确地分类，那么它的权重就得到提高。通过这样的方式，AdaBoost 方法能“集中聚焦”学习那些难区分（更富信息）的样本，最终得到分类效果理想的强分类器。



扩展：AdaBoost 为每个分类器都分配了一个权重值 α ，这些 α 值是基于每个弱分类器的错误率进行计算的，其中错误率 ε 的定义为：

$$\varepsilon = \frac{\text{未正确分类的样本数目}}{\text{所有样本数目}}$$

α 的计算公式如下：

$$\alpha = \frac{1}{2} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$$

为每个弱分类器计算出 α 后，就可以对样本权重向量 \mathbf{D} 进行更新，以使得那些正确分类的样本的权重降低而错误分类的样本的权重升高。样本权重向量 \mathbf{D} 的计算方法如下。

如果某个样本被正确分类，则该样本的权重被调整为：

$$D_i^{(t+1)} = \frac{D_i^{(t)} e^{-\alpha}}{\text{Sum}(\mathbf{D})}$$

反之，如果某个样本被错误分类，则该样本的权重被调整为：

$$D_i^{(t+1)} = \frac{D_i^{(t)} e^{\alpha}}{\text{Sum}(\mathbf{D})}$$

在现实的人脸检测中，只靠一个强分类器还是难以保证检测的正确率，所以我们通常训练有多个强分类器。这些强分类器虽然性能优越，但因为构造比较复杂，所以计算起来比较费时。如果同时计算它们的话，速度肯定不快。怎么办？可通过设置层层关卡的方式，一关一关地过，直到通关！也即，我们可通过筛选式**级联（Cascaded）**的方式来大大地提高速度。具体来说，级联结构的分类器是由一系列强分类器串联组成。对要识别的样本进行判别时，只有被前面一级的分类器判别为正的样本（“是人脸”）才被送入后面的分类器继续处理，反之则被认为是负样本（“非人脸”）而直接踢出局。最后，只有被所有分类器都判别为正的样本才会被输出。在设计级联结构时，前面几级的分类器都选用结构比较简单（计算快速）的，使用的特征数也较少，但检测速率很高，可以一下子就滤除那些与目标差异较大的负样本；后面级次的分类器则使用更多的特征和更复杂的结构（计算费时），这样可以细致地滤除那些与目标相似的负样本。这样做的一个重要好处是：可以将大量的不含人脸的子窗口图像“阻挡”在级联分类器的最前面几层，即只需简单计算就被直接滤除，而不进入后面的阶段，因而大大地减少了计算量。

如图 6-68 所示，举个最简单的**决策树（Decision Tree）**例子，实际级联结构就类似于这么一种退化的决策树。假设我们使用 3 个 Haar-like 特征 f_1 、 f_2 、 f_3 来判断输入数据是否为人脸，可以建立如下决策树进行判别。可以看出，这种判别方式非常简单、快速，只需做 if-then 条件规则判断即可。

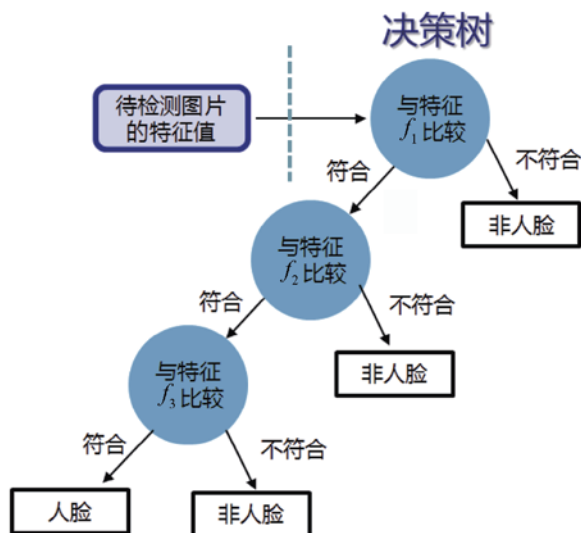


图 6-68 筛选式级联的工作原理



提示：与前面 6.4 节的感知机和 SVM 相比，决策树（如分类回归树 Classification And Regression Trees, **CART**）以及**随机森林（Random Forest）**这种对复杂数据不断进行子划分的方式很适合于非线性问题，因为当数据拥有众多特征且特征之间关系十分复杂时，要对所有的非线性数据构建一个统一的非线性划分函数将变得非常困难。但在决策树的生成过程中，往往需要对树进行剪枝简化，以避免构建出**过于复杂和有噪声**的决策树，导致出现**过拟合**现象（**Over-fitting**，即对**已知的训练数据**的分类过于准确，连细枝末节上的“坏分子”噪声都好坏不分地拟合进去了，导致对**未知的测试数据**的分类反而不准确了、降低了对未知数据的预测能力，即**泛化能力 Generalization Ability**），正所谓“过犹不及”。

下面我们展示如何用 OpenCV 来编写人脸检测代码。注意：OpenCV 已训练好了 Haar 分类器，所以我们无须进行上面的分类器训练过程，只需应用级联分类器进行判别即可。

程序 6-2：从文件中读取一幅图像并检测出其中的人脸

```

#include "cv.h"
#include "highgui.h"

// 基于 Haar 特征的人脸检测要用到的级联分类器文件
string face_cascade_name = "haarcascade_frontalface_alt.xml" ;
// 级联分类器类
CascadeClassifier face_cascade;
string window_name = "result" ;

void detectAndDisplay( Mat frame );

int main( int argc, char** argv ){
    Mat image;

```

```

    image = imread( argv[1] );

    if( argc != 2 || !image.data ){
        printf( "[error] 没有图片\n" );
        return -1;
    }

    // 加载人脸检测所用的分类器，并判断是否载入成功，如果不成功则提示
    if( !face_cascade.load( face_cascade_name ) ){
        printf( "[error] 无法加载级联分类器文件! \n" );
        return -1;
    }

    detectAndDisplay(image);

    waitKey(0);
}

void detectAndDisplay( Mat frame ){
    // 函数使用针对某目标物体训练的级联分类器，在图像中找到包含目标物体的矩形区域，并且将这
    // 些区域作为一序列的矩形框返回，最终检测结果保存在 Rect 变量中
    std::vector<Rect> faces;
    Mat frame_gray;

    // 由于大部分的脸部检测算法对光照、脸部大小、位置表情等敏感，所以一般需利用 cvtColor 函
    // 数将其转化为灰度图像，并利用 equalizeHist 函数进行直方图（Histogram）均衡化来增强
    // 图像的对比度。在本例中，将变量 frame 转换成灰度图，并输出到变量 frame_gray
    cvtColor( frame, frame_gray, CV_BGR2GRAY );
    equalizeHist( frame_gray, frame_gray );

    // 对图片 frame 进行识别检测
    face_cascade.detectMultiScale( frame_gray, faces, 1.1, 2, 0|CV_HAAR_
SCALE_IMAGE, Size(30, 30) );

    // 人脸检测结果用红圈框出
    for( int i = 0; i < faces.size(); i++ ){
        Point center( faces[i].x + faces[i].width*0.5, faces[i].y +
faces[i].height*0.5 );
        ellipse( frame, center, Size( faces[i].width*0.5, faces[i].
height*0.5), 0, 0, 360, Scalar( 255, 0, 255 ), 4, 8, 0 );
    }

    imshow( window_name, frame );
}

```

以上代码的人脸检测效果如图 6-69 所示，用一个红色圆框把人脸定位出来了。请仔细数一下，刨去那么多的注释代码，真正的代码是不是也只寥寥数行而已，但却可实现这么智能化的效果。



图 6-69 OpenCV 的人脸检测定位效果（图片为笔者本人）

附：OpenCV 的主要功能

- 图像数据操作（内存分配与释放，图像复制、设定和转换）
- 图像 / 视频的输入 / 输出（支持文件或摄像头的输入，图像 / 视频文件的输出）
- 矩阵 / 向量数据操作及线性代数运算（矩阵乘积、矩阵方程求解、特征值、奇异值分解）
- 支持多种动态数据结构（链表、队列、数据集、树、图）
- 基本图像处理（去噪、边缘检测、角点检测、采样与插值、色彩变换、形态学处理、直方图、图像金字塔结构）
- 结构分析（连通域 / 分支、轮廓处理、距离转换、图像矩、模板匹配、霍夫变换、多项式逼近、曲线拟合、椭圆拟合、狄劳尼三角化）
- 摄像头定标（寻找和跟踪定标模式、参数定标、基本矩阵估计、单应矩阵估计、立体视觉匹配）
- 运动分析（光流、动作分割、目标跟踪）
- 目标识别（特征方法、HMM 模型）
- 基本的 GUI（显示图像 / 视频、键盘 / 鼠标操作、滑动条）
- 图像标注（直线、曲线、多边形、文本标注）

6.11.2 OpenGL 与 3D 图形绘制

6.11.1 节介绍了计算机视觉领域的开源库 OpenCV，本节接着介绍一下计算机图形学领域的开放库 OpenGL。OpenGL 的英文全称是“Open Graphics Library”，顾名思义，OpenGL 是“开放的图形程序库”。虽然 DirectX 在家用市场全面领先，但在专业高端绘图领域，OpenGL 是不能被取代的主角。OpenGL 定义了一个跨平台的 3D 图形编程接口，具有很好的移植性，广泛应用于 CAD 设计、游戏开发、制造业及虚拟现实等行业领域中。

下面，我们介绍如何使用 OpenGL 显示一个 3D 球体，代码如下所示。

程序 6-3：使用 OpenGL 显示一个 3D 球体

```
#include <GL/glut.h>           // 包含 OpenGL 实用库
#include <stdlib.h>
#include <math.h>

// 声明一个二次曲面对象，如球体
GLUQuadricObj *quadObj;

// 设置光照参数，如环境光、漫射光以及光源位置
static float light_ambient[] = {0.1,0.1,0.1,1.0};
static float light_diffuse[] = {0.5,1.0,1.0,1.0};
static float light_position[] = {90.0,90.0,150.0,0.0};

void myInit(void)
{
    glClearColor(0.0,1.0,0.0,0.0);           // 设置背景色为绿色
    glEnable(GL_DEPTH_TEST);                 // 启用深度测试

    // 设置光照
    glLightfv(GL_LIGHT0,GL_AMBIENT,light_ambient);
    glLightfv(GL_LIGHT0,GL_DIFFUSE,light_diffuse);
    glLightfv(GL_LIGHT0,GL_POSITION,light_position);

    // 激活光照
    glEnable(GL_LIGHTING);                   // 打开光照
    glEnable(GL_LIGHT0);                     // 打开光源 0
    glShadeModel(GL_SMOOTH);                 // 启用阴影平滑
}

void myDisplay(void)
{
    glClear(GL_COLOR_BUFFER_BIT|GL_DEPTH_BUFFER_BIT); // 清除屏幕和深度缓存

    // 创建一个二次曲面物体，如球体
    quadObj = gluNewQuadric();

    // 将这个球体绘制出来，半径为 3，精细程度为 20
    glPushMatrix();
    gluSphere(quadObj,3.0,20.0,20.0);
    glPopMatrix();

    // 将二次曲面对象删除
    gluDeleteQuadric(quadObj);

    glFlush();                               // 强制刷新缓冲
```

```

}

void myReshape(int w,int h)
{
    glViewport(0,0,(GLsizei)w,(GLsizei)h);    // 设置视区尺寸
    glMatrixMode(GL_PROJECTION);                // 选择投影矩阵
    glLoadIdentity();                            // 重置投影矩阵
    gluPerspective(45.0,(GLfloat)w/(GLfloat)h,1.0,50.0); // 指定透视投影的观察空间
}

```



提示：透视投影类似于人眼观察真实世界的视觉效果，即离视点近的物体大，离视点远的物体小，远到极点即为消失，成为灭点。比如我们看长长的铁轨（如图 6-70 左边所示），原本平行的左右两边似乎要在无穷远的地方相交。而如图 6-70 右边所示，透视投影可用一个视锥体模型来描述，OpenGL 中函数 `gluPerspective(GLdouble fovy, GLdouble aspect, GLdouble zNear, GLdouble zFar)` 中的各参数意义分别为：`fovy`（视角）、`aspect`（宽/高比）、`zNear`（Z 轴近平面距离）、`zFar`（Z 轴远平面距离）。注意：只有当 3D 模型处于由 Z 轴近平面和 Z 轴远平面围成的视野范围内时，才会被 OpenGL 显示在二维屏幕上。

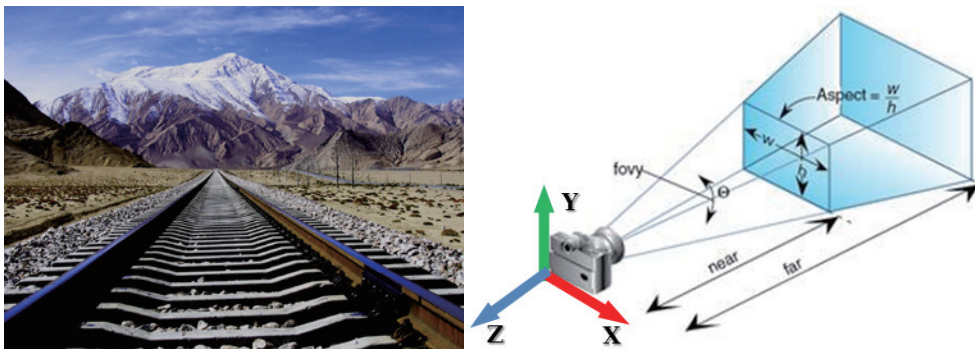


图 6-70 透视投影的示意图（图片来源：xizanglvxing）

```

glMatrixMode(GL_MODELVIEW);    // 选择模型观察矩阵
glLoadIdentity();              // 重置模型观察矩阵
glTranslatef(0.0,0.0,-15.0);   // 将图形沿 z 轴负向移动（移入屏幕 15 个单位），
                                // 以便出现在可观察空间内（具体可参见图 6-70）
}

int main(int argc,char ** argv)
{
    // 初始化
    glutInit(&argc,argv);
    glutInitDisplayMode(GLUT_SINGLE|GLUT_RGB|GLUT_DEPTH);
    glutInitWindowSize(400,400);

    // 创建窗口
    glutCreateWindow("绘制一个 3D 球体");
}

```

```
// 绘制与显示
myInit();
glutReshapeFunc(myReshape);
glutDisplayFunc(myDisplay);

glutMainLoop();
return 0;
}
```

以上代码编译运行后的效果如图 6-71 所示。上面的代码实际上也没有几行，却实现了诸如二次曲面球体生成、光照、视角放置等诸多功能，是不是特别方便？

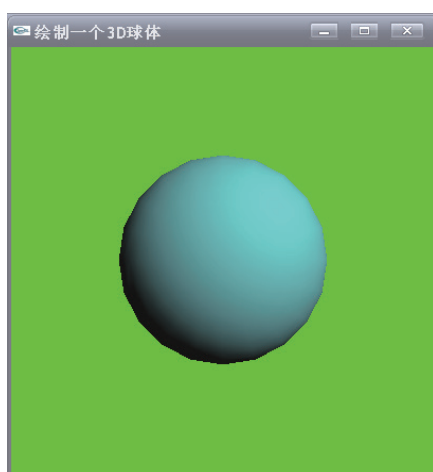


图 6-71 OpenGL 生成和显示的球体

附：OpenGL 的主要功能

- 建模：OpenGL 图形库除了提供基本的点、线、多边形的绘制函数外，还提供了复杂的三维物体（球、锥、多面体、茶壶等）以及复杂曲线和曲面绘制函数。
- 变换：OpenGL 图形库的变换包括基本变换和投影变换。基本变换有平移、旋转、缩放、镜像 4 种变换，投影变换有平行投影（又称正射投影）和透视投影两种变换。其变换方法有利于减少算法的运行时间，提高三维图形的显示速度。
- 颜色模式设置：OpenGL 颜色模式有两种，即 RGBA 模式和颜色索引（Color Index）。程序员还可以选择平面着色和光滑着色两种着色方式对整个三维景观进行着色。
- 光照和材质设置：用 OpenGL 绘制的三维模型必须加上光照才能与客观物体更加相似。OpenGL 提供了管理 4 种光（辐射光、环境光、镜面光和漫反射光）的方法，另外，还可以指定模型表面的反射特性。
- 纹理映射（Texture Mapping）：OpenGL 提供的一系列纹理映射函数使得开发者可以十分方便地把真实图像贴到景物的多边形上，从而在视窗内绘制逼真的三维景观。
- 位图显示和图像增强功能：除了基本的复制和像素读写外，还提供融合（Blending）、抗锯齿（也被称为：反走样，Antialiasing）和雾（Fog）的特殊图像效果处理。以上 3 条可

使被仿真物更具真实感，增强图形显示的效果。

- 双缓存动画（Double Buffering）。为了获得平滑的动画效果，需要先在内存中生成下一幅图像，然后把已经生成的图像从内存复制到屏幕上，这就是 OpenGL 的双缓存技术（Double Buffer）。

此外，利用 OpenGL 还能实现深度暗示（Depth Cue）、运动模糊（Motion Blur）等特殊效果，从而实现了消隐算法。

第7章

创客：个人3D打印机的创造者

《晏子春秋·内杂下》有云：“为者常成，行者常至”。大意是：努力去做的人常常可以成功，不倦前行的人常常可以达到目的地。这句话若用在个人 3D 打印机的发展历程上，也同样精辟。实际上，正是由于**创客（Maker）**们的努力，才使得 3D 打印机这种价格原本高高在上的工业设备，走进了普通家庭和日常生活。**创客的特点之一就是要亲自动手做（Make）**；特点之二是把做好的新奇玩意儿拿出来分享，而且是那种把设计图纸、源代码拿出来与大家彻彻底底共享；特点之三是利用网络沟通的优势，和志同道合的人一起合作，共同探讨和完善产品的创意定位、技术难点、融资方式、赢利模式、营销手段等。

目前创客文化已经在国外如火如荼，一大堆新奇的技术和产品随之诞生。加上开源共享的催化作用，进一步使很多新概念和新产品遍地开花。所以将创客称之为第三次工业革命的启蒙者一点儿也不为过。正是他们的热情和无私，才打破了大工厂、大公司和跨国企业的技术封锁，为“个人智造”、“家庭智造”、“网络社区智造”新时代的到来打下了坚实的基础。

7.1 创客文化与开源DIY

“**创客**”一词来源于英文单词“Maker”，是指喜欢动手制作，努力把各种创意转变为现实产品的人。创客在国内才刚刚萌芽，而在国外发展历史却比较长。乔布斯堪称是创客界的元老，从小爱“倒腾”的他曾尝试改变电话中的脉冲频率来打免费电话，也正因为具有对创新的执著追求精神才成就了今天的“苹果”。当然，本书长篇讨论的、现已风靡全球的 3D 打印机也是出自创客之手。

创客是一群喜欢或者享受创新的人，追求自身创意的 DIY（Do It Yourself，自己动手）实现。创客不分行业或科技含量的高低，只要愿意去实现自己创意和想法的人，都是创客。也正因为如此，创客文化中的多元与分享使其极具包容性和亲和力，能让更多人参与进来。麻省理工学院（MIT）电子工程和计算机专业的辛普森·星（Star Simpson）是一个开源硬件的沉溺者，她说：“所以在我看来，低科技和高科技并无区别，它们都是把我们的设想变成现实的工具”。

如果要把创客的起源解释清楚，那么我们不得不提“**黑客（Hacker）**”。50 多年前 MIT 的一

些学生可谓是最早的黑客了，他们业余喜欢鼓捣一些计算机编程，且技术实力很强，也就是现在所说的“老鸟”或者“大牛”。可以简单概括为一句话：要想称得上“黑客”，他必须有创新、有风格、有技术含量（摘自 Steven Levy 的《黑客：计算机革命的英雄》）。黑客伦理和黑客文化紧密关联，其核心精神是“开放、共享、分权和对技术的崇拜”，并最终发起了 GNU 计划和自由软件运动。随着自由软件的传播发展，促生了带有商业化倾向的开放源代码软件，开源软件的诞生进一步拓展了自由软件的发展道路，迎来了开源大发展的 21 世纪。既然软件已经开源了，那么硬件呢？

由此，开源硬件也就呼之欲出了！开源硬件特别是以 Arduino 为首的硬件产品催生了一个新的群体——创客。而开源硬件和创客群体也借着互联网，仅用 5 年时间就快步发展成熟。

由上面的分析可以看出，创客群体的产生得益于开源硬件，而开源硬件继承自开源软件，开源软件脱胎于自由软件，而自由软件则凝聚了“黑客文化”。所以简单来说，“创客文化”的本源与实质就是“黑客文化”。

信息时代与民间创造不无关系，比如第一台苹果电脑 Apple I 就是苹果公司的另一个创始人史蒂夫·沃兹尼亚克（Steve Wozniak）在 1976 年的“家酿计算机俱乐部”中制作的。几十年后随着互联网和开源运动的发展，越来越多的非专业人士也拥有了多学科跨界创造的知识，创客文化得到了迅速的发展。创客的共同特点是创新、实践与分享，但这并不意味着他们都是一个模子里铸出来的人；相反，他们有着丰富多彩的兴趣爱好，以及各不相同的特长，一旦他们聚到一起，相互协调，发挥自己特长时，就会爆发出巨大的创新活力。创客大体可分为如下几类。

创意者

他们是创客中的精灵，他们善于发现问题，并找到改进的办法，将其整理归纳为创意和点子，从而不断创造出新的需求。

设计者

他们是创客中的魔法师，他们可以将一切创意和点子转化为详细可执行的图纸或计划。

实施者

他们是创客中的剑客，没有他们强有力的行动，一切都只是虚幻泡影，而他们高超的剑术，往往一击必中，达成目标。

随着自由文化的不断发展，现在已出现了各种“客”，如黑客、极客、威客、创客等，让人眼花缭乱，现统一比较如下。

黑客（Hacker）：特指那些热爱软件编程，善于改造计算机的“黑客”，而不是在网络上盗取个人信息或者破坏网络的“Cracker（骇客、破解者）”。

极客（Geek）：原先特指那些研究计算机和网络技术到痴迷的“宅男”，因此颇有些神秘气息。而近年来随着“宅男”文化的普及，也可以代指一种为了好玩而仿技术流的个性化时尚现象。极客和时尚的关系非常密切，比如制造非常酷的电子设备，为普通的时装增加非常酷的 LED 灯装饰，制作一个真实版的变形金刚等。极客拥有很多特性，酷是最本质的属性，一切都是为了酷这个目标。有时可以为了技术而技术，也就是为了“秀”。技术未必很高超，但效果往往特别

吸引眼球。

威客：威客的英文 Witkey 是由 Wit（智慧）、key（钥匙）两个单词组成的，也是 The key of wisdom 的缩写，是指那些通过网络平台将自己的知识、技能、专长、经验等出售给别人并且换取经济收益的人。一般文稿工作者、广告设计师、程序员所占的比例较大，因为这类工作一般比较短期，按件完成，容易整体外包，而且工作时间灵活。

Maker 表示“**创客**”，伴随着开源硬件兴起的一个名词。然而，硬件电路板只是个躯壳，一个产品的灵魂在于软件与控制，所以创客还必须像黑客那样精通软件编程。同时，为了让自己的作品显得与众不同，创客也要像极客那样让自己的作品酷。最后，光凭兴趣肯定难以养家糊口，所以创客在产品的最后阶段也需要像威客那样学会思考赢利模式和营销手段。

因此，在本书作者看来，创客是所有“客”发展的最新阶段，也可以说是集大成者，因为“创客”集各种“客”的特点于一身，也即创客（Maker）= 黑客（Hacker）+ 极客（Geek）+ 威客（Witkey）。也好，“创客”正如金庸老先生在《笑傲江湖》中所说的：“文成武德，英明神武。千秋万载，一统江湖”，也许各种“客”的烦乱混杂目前终于可以消停一段时日了。

7.2 五花八门的创客杰作：从玩具到高速跑车

也许你还是觉得创客很神秘，缺乏一个直观的了解。我记得有一句话是这么说的：了解一个人，要看他所做的事，正所谓“观其行”。下面就来给大家展示一下创客们的一些作品，看看他们喜欢做哪些事。

如图 7-1 所示，这是一辆基于光电感应与雷达控制的现代投石车玩具，涉及的专业知识有光学、机械、电子控制等。有了 3D 打印机，你将来几乎不用再去买专门的玩具了，因为网上有成千上万的新奇玩具模型供免费下载，直接打印出来就可以玩。目前比较火的个性化玩具定制网站有 MakieLab，所推出的第一款 3D 打印玩具 Makies 已达到了欧盟玩具安全标准。此外，还有一个叫 MAQET 的网站，用户定制完成后，只需单击“Make it Real”按钮，几天后打印好的玩具就可送货上门。

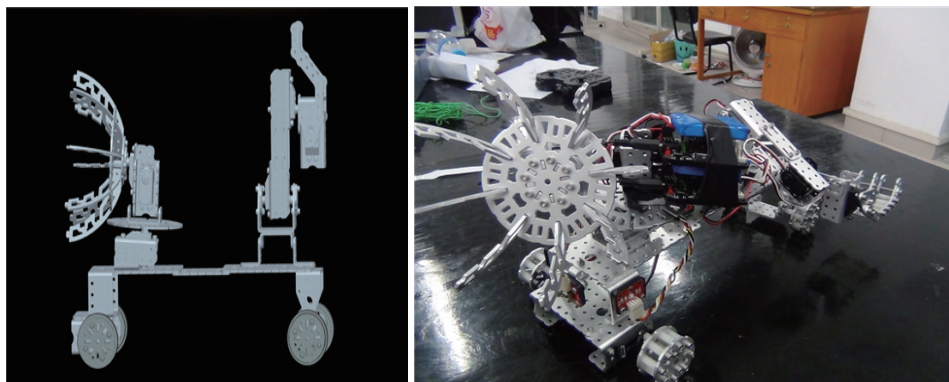


图 7-1 基于光电感应与雷达控制的现代投石车（图片来源：内蒙古工业大学）

如图 7-2 所示，这是一件超逼真的煎鸡蛋创意文化衫。说到煎鸡蛋你见过多大的？蛋黄有多大？今天让大家开开眼。瞧瞧图中的煎鸡蛋吧。大吗？酷吗？



图 7-2 超逼真的煎鸡蛋创意文化衫（图片来源：nixon）

如图 7-3 所示，这款 Giuseppe 概念赛车是设计师 Jaemin Park 的作品，它的出现打破了传统的赛车设计风格，每一个设计细节都考虑到了如何减少空气阻力。还有可随意调整宽度的独立前翼，增加了车体的灵活性，设计大大提高了赛车应对直道和弯道的能力。

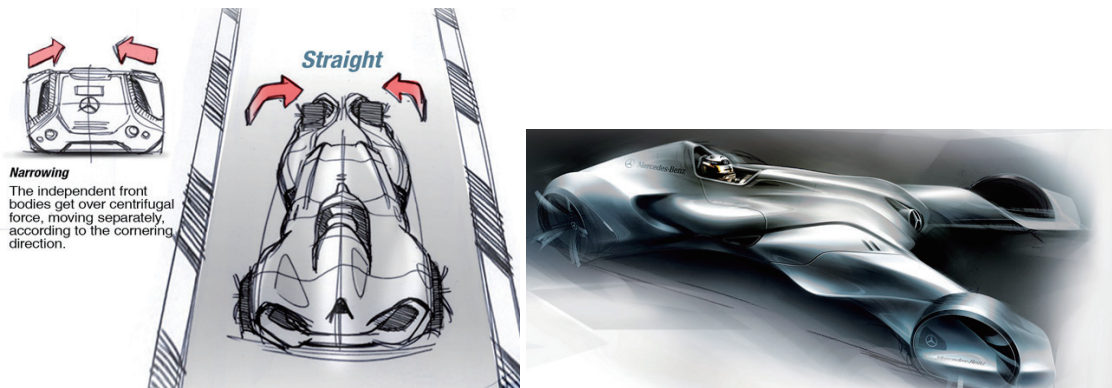


图 7-3 Giuseppe 概念赛车（图片来源：Jaemin Park）

Whisk 是款多才多艺的打蛋器，是设计师 Ivan Zhang 的创意，如图 7-4 所示。它的结构非常巧妙，让打蛋器适当地内凹，变成一个能放一个鸡蛋的凹槽。这种设计可以让蛋清漏下，蛋黄留在上面，从而实现蛋清和蛋黄的分离。

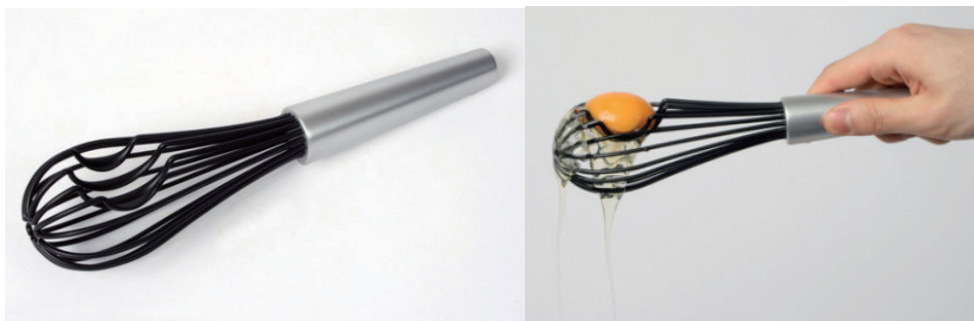


图 7-4 Whisk：可分离蛋清和蛋黄的打蛋器

Oculus Rift 穿戴式 3D 眼镜——让你所见的，就是真实的。戴上 Oculus Rift，你甚至会有一种错觉：你就在游戏里，你想和前面的人说话、你抬头可以看到天上在下雪、你想伸手去摸雪花、你看见有人迎面走过来下意识想要躲闪。戴上 Oculus Rift，拿一把 Razer Hydra 改装的枪，你就可以在自己的家里体验使命召唤里的真实战场，如图 7-5 所示。



图 7-5 Oculus Rift 穿戴式 3D 眼镜

还认为触摸屏是很酷的技术吗？如今已经可以手指不接触屏幕来实现触摸的效果了。Leap Motion 拥有精细到毫米级的感应、流畅到极致的操控体验，比如用户不用触摸屏幕就可以隔空玩“切水果”游戏，如图 7-6 所示。

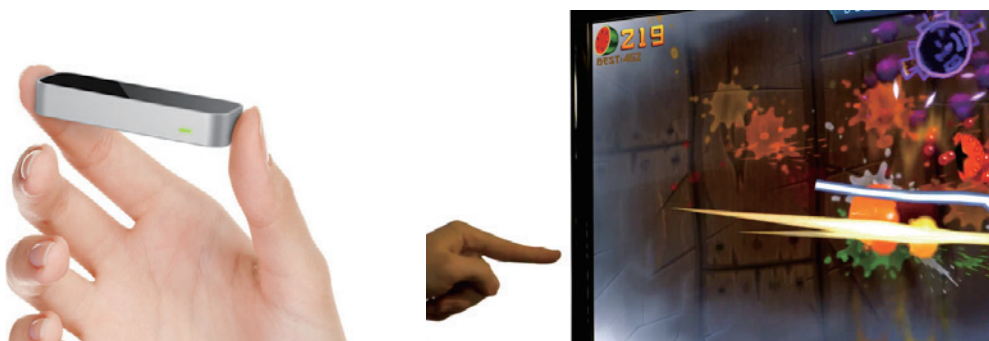


图 7-6 Leap Motion 隔空玩“切水果”游戏

最近更酷的设备就是如图 7-7 所示的这个由 MYO 推出的手臂环了。这个环套在人的胳膊上，

可以检测人手臂肌肉的运动，并且将它们转化为电子信号，结合其他传感信息，就可以捕捉到人的复杂的手势动作，以此来判断人们想要做什么，如射击游戏中的举枪动作、握拳、控制音乐播放器等。



图 7-7 MYO 推出的手臂环

万向跑步机由 Virtuix 公司推出，产品名叫 Omni，如图 7-8 所示。它会将人的方位、速率和里程数据全部记录下来并传输到游戏当中，玩家能够在现实中 360° 地控制游戏虚拟角色的行走和运动（还有，可结合前面的 Oculus Rift 眼镜和一把 Razer Hydra 改装的枪）。



图 7-8 Omni 万向跑步机

在游戏中感受不到任何身体的撞击或子弹的冲击，这显然是不合理的。而穿上 ARAIG 盔甲，如图 7-9 所示，如果游戏中人物给你腰部一记重拳，你的身体就能真真实实地感受到腰部受到一个强大的压力（当然仍在安全范围内），如果一颗子弹打过来，你也能感受到像穿了防弹衣后子弹的冲击，从而使玩家的游戏体验变得更真实。其他如雨滴、轻拍、撞击、跌打等动作，都可以通过这套设备模拟出来，使游戏效果更加真实。遗憾的是，这个 ARAIG 项目目前并没有在 Kickstarter 众筹网站筹集到足够的资金。

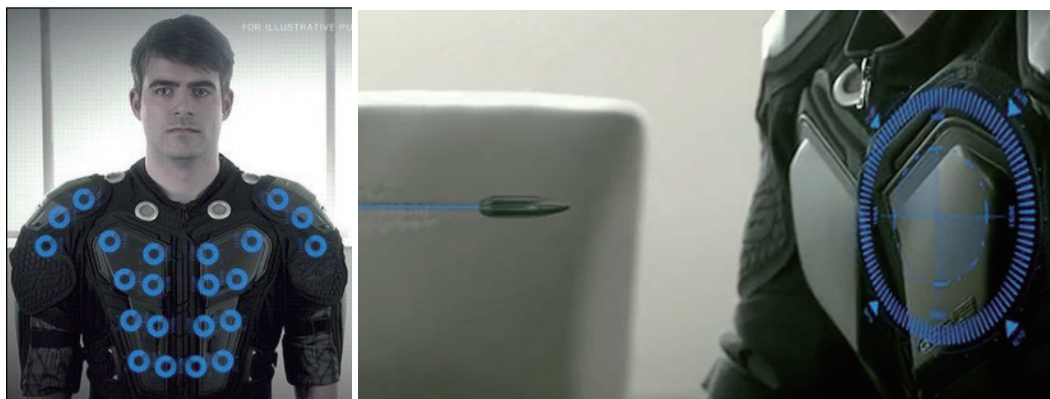


图 7-9 ARAIG 盔甲

看了上面的介绍，相信你对创客已经非常神往了吧？一台 3D 打印机放在书桌旁，伴着悠扬的音乐，你在计算机上设计出来的数字化模型，通过几个小时的打印，就能以实物形式拿到你手中，你再安装几个电控部件，一个精致的恐龙模型就摇头摆尾地行走起来了。这样的创作模式是否激发了你久违了的生活热情呢？

7.3 寓教于乐：3D打印出你的个人数学博物馆

3D 打印除了在产品设计开发、艺术、食品、医疗、个人制作等方面的应用之外，在教育上还有着特别的应用。以数学为例，估计本书的很多读者都会觉得数学枯燥无味。而通过 3D 打印技术，却可以让原本枯燥的数学公式变得直观有趣。

下面是国外一名创客给大家隆重推荐的个人数学博物馆！在 3D 数学博物馆里，你不仅可以欣赏各种叫得出名与叫不出名来的三维立体几何图形，甚至连神秘莫测的四维图形也可以打印出来了，即直观地显示四维图形在三维世界的投影。值得注意的是，对于这些复杂中空的结构，之前的任何切削工艺都是不可能加工的，唯有 3D 打印出现后才得以实现。

我们首先从球体开始，该球体由许多个近似菱形组成，如图 7-10 所示。



图 7-10 3D 打印的球体（图片来源：George W. Hart）

如图 7-11 所示是 Sierpinski 分形四面体，你在第 4 章 4.2.1 节的分形中应该见过。

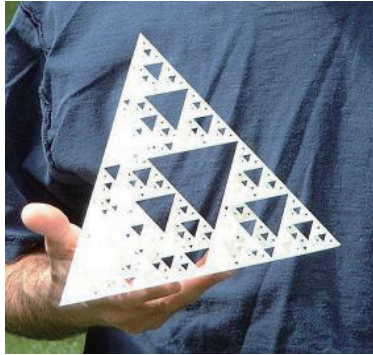


图 7-11 Sierpinski 分形四面体

图 7-12 中是另一个著名的分形：孟结海绵（Menger Sponge），是一个三级分形，也就是说这个分形中有 3 种不同大小的孔。该分形的有趣之处在于：它的表面积会随着它级数的增长以指数方式增长。

如图 7-13 所示，这是一个双曲面模型。

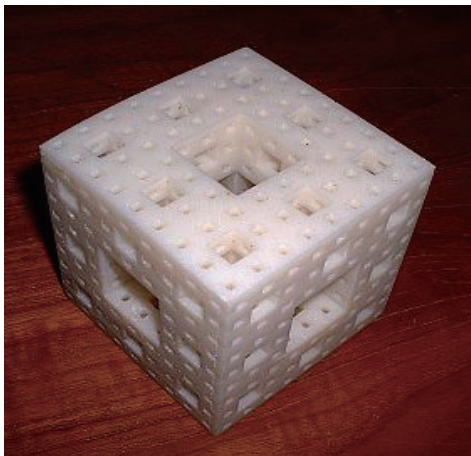


图 7-12 孟结海绵（Menger Sponge）

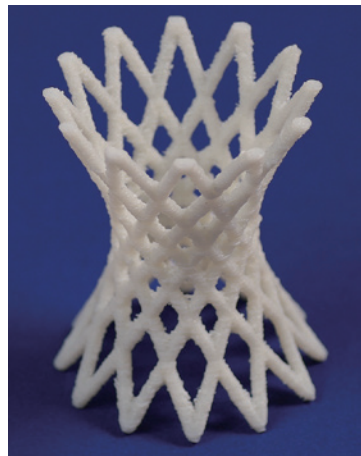


图 7-13 双曲面模型

如图 7-14 所示，这些可认为是螺旋环面结构的各种衍生变种。

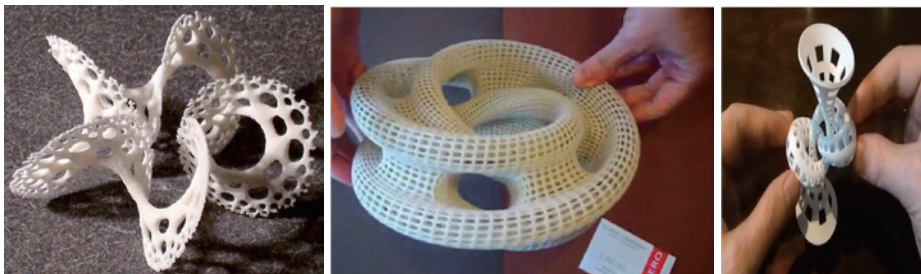


图 7-14 螺旋环面结构的各种衍生变种

我们知道，由一个四维物体，可以计算出其在三维空间的“投影”。这个投影往往是个繁复而美丽的三维物体。与传统工艺相比，3D 打印具有制作复杂模型的巨大优势，可以很容易地制作出这些似乎只存于思想中的结构（因为四维根本无法画出来）。

图 7-15 中是一个 120 胞体（120-Cells），是由 120 个正十二面体组成的四维结构“投影”而成的。该四维结构原本由一个大正十二面体被 119 个小正十二面体填充组成。但是在投影到三维空间时，除了最外层和最内层的两个十二面体还是正十二面体之外，其他的十二面体的角度都产生了相应的投射扭曲。

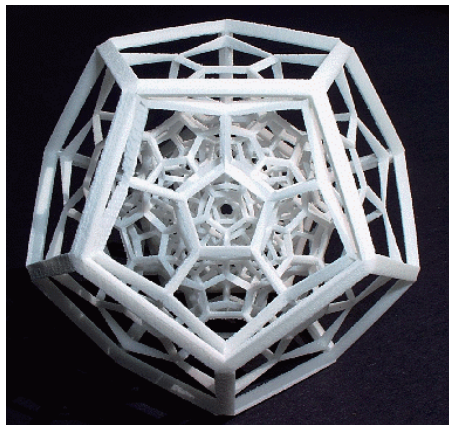


图 7-15 120 胞体（120-Cells）



提示：世界上最早意识到并理解多维空间存在的人，可追溯到 19 世纪瑞士几何学家 Ludwig Schläfli（路德维希·施莱夫利），他为此研究出相应的多维几何学。如图 7-16 所示，点的维度是 0，线的维度是 1（ X 坐标轴），面的维度是 2（ X/Y 坐标轴），体的维度是 3（ $X/Y/Z$ 坐标轴），如果再加一个 W 坐标轴，就进入到四维空间了。（实际上，我们在三维空间看不见四维的 W 坐标轴，图中只是一个假想。）

我们先分析一下二维（Dimension）的情况。同理，2D 中的平面人看不见 3D 空间的我们，而且看 2D 空间本身的那个正方形每次也只能看到一条线段，更无法“穿透”进正方形内部看见那位美女。而在 3D 中的我们，不仅一下就能看出是正方形，而且还能看到正方形里面的那位美女。也即，在低维空间里原本属于内部的东西，到了高维空间都会变成外部的，形象地说：低维空间不过是高维空间的表皮。

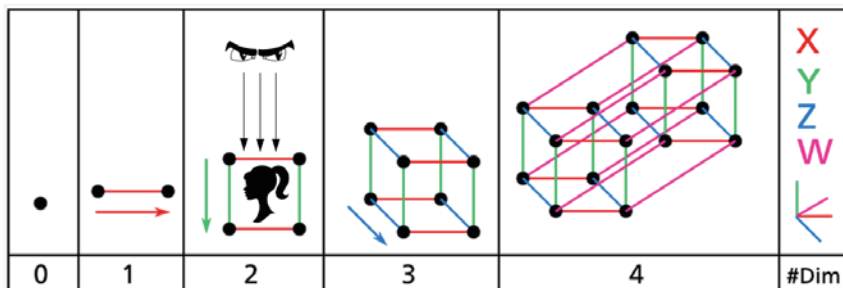


图 7-16 从一维到四维空间的性质图解（图片来源：维基百科）

类似地，处在 3D 空间中的我们，也看不见 4D 空间的物体，而且看 3D 空间本身的物体每次也只能看到一个面，无法穿透到内部，比如别人身体里心脏的跳动。而在 4D 空间中的“四维人”，却可以一下子（同一时间）就透视我们身体内外的每一个细节。

既然我们看不见 4D 空间中的物体，那么如何感受它呢？将它投影到 3D！正如拿着一个 3D 的球穿过一张 2D 的平面，平面上的 2D 人会发现一开始出现了一个点，然后扩大成一个圆，接着再慢慢缩小成一个点直到消失。类似地，一个 4D 的球体穿过我们 3D 空间，也会从一个小球渐渐变大到一个大球，再渐渐变小直到消失。实际上，4D 或更高维的空间形状性质我们一般都可通过这种低维类推的方式来考察。如图 7-17 所示，0-单纯形就是点（0 维）、1-单纯形就是线段（1 维）、2-单纯形就是三角形（2 维）、3-单纯形就是四面体（3 维），因此我们可以类推出 4 维空间中的 4-单纯形将是一个有着 5 个顶点（以及 10 条线段、10 个三角形面、5 个四面体）的五胞体。这里的单纯形（Simplex，也被译作单形）是代数拓扑中的基本概念，是三角形和四面体的一种高维泛化，一个 n 维单纯形是指包含 $n+1$ 个节点的凸多面体，即 $n+1$ 个仿射无关的点的集合的凸包。

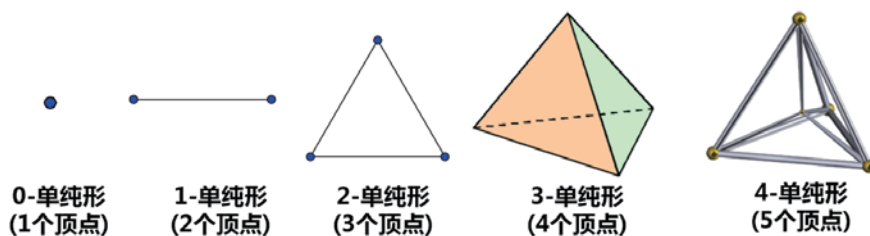


图 7-17 利用从低维类推的方法来分析高维空间形状的性质

如图 7-18 所示是一个 4D 超正方体绕平面旋转时在 3D 上的透视投影变化，正如我们拿一个 3D 正方体旋转不同角度投影到 2D 纸面上时，2D 投影形状也会往复变化。

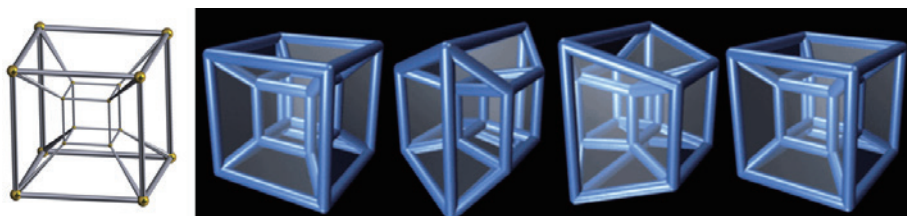


图 7-18 四维超正方体绕平面旋转时在三维上的透视投影变化

还有一个有趣的现象是，那位 2D 平面上的美女被 2D 正方形的四条边所困住，而位于 3D 中的我们只要把她往第 3 维 Z 轴一抬，她就解围了。同理，处在 3D 空间中的我们若被困在一个正方体里面，4D 空间的人只需把我们往第 4 维 W 轴一抬，我们就解围了。

如图 7-19 左边所示，莫比乌斯带（Möbius band）是一种拓扑学结构，它没有正反面之分。想象一张长条纸，把它扭转一圈后首尾相连，就会发现原来的一面与其反面相连。如果你在这个纸面上沿着一个方向走，不用翻栏，就能够走遍这张纸条的所有面并回到起点。莫比乌斯环

看着像一个 2D 结构，但是它本身却只能在 3D 空间存在。在图 7-19 的右图中，3D 打印出来的**克莱因瓶（Klein bottle）**类似于莫比乌斯带，也是一种不可定向的闭曲面，没有“内部”和“外部”之分。在这个奇怪的管状物里行走，你能经历空间的正面和反面，并回到起点。其实真正的克莱因瓶存在于 4D 空间：克莱因瓶看上去好像有一个与自己相交的部分，然而在 4D 空间它并不相交，就像莫比乌斯环在 3D 空间不相交一样。

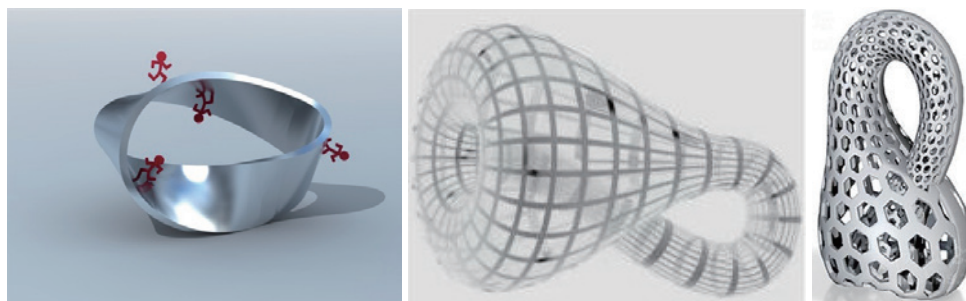


图 7-19 莫比乌斯带和克莱因瓶（图片来源：Shapeways）

假如我们的宇宙也是一个由扭曲的空间形成的克莱因瓶结构，那么这个宇宙虽然大小有限，但是却并没有边界，你沿同一个方向走，会永远地不停循环下去。

看完之后，有些对数学天生不感兴趣的读者在惊叹之余，可能还是觉得没有直观感受，比如看不见内部。没问题，我们可以把这些数学 3D 模型用巧克力食品打印机打出来，然后你就可以一边品尝，一边欣赏内部的优美几何结构了。

7.4 创客之开源硬件Arduino（阿德伟诺）

“不要重新发明轮子”。在全球科技更新换代周期不断缩短的背景下，这句话所蕴含的道理显得尤为深刻。对于电子设计工程师以及爱好者、艺术家来说，Arduino 开源硬件无疑是一个能够帮助他们快速实现设计梦想的工具箱，只需站在前人的肩膀上，而不必从无到有地发明这些工具和组件。本章将循序渐进地介绍 Arduino 的背景、硬件、开发环境以及一个简单、完整的开发项目，为我们的进一步开发奠定坚实基础。

几乎任何人，即使不懂计算机编程，也能用 Arduino 做出很酷的东西，比如对感应器做出回应，闪烁灯光，还能控制马达，Arduino 的存在让制作硬件的门槛极大降低，热爱动手的人们不需要太高的成本就能创造出好玩的硬件产品。由于 Arduino 具有高度模块化的特点，所以有时也形象地叫它“电子积木”。

7.4.1 Arduino 简介

开源硬件作为一个附件或设备，允许任何人随意复制或修改硬件设计图纸。你可以自己下载规格说明书后组装一台，或者从制造商那里购买并支付一小部分的组装费。

Arduino 由 5 个国际工程师研发，他们分别是 Massimo Banzi、Gianluca Martino（意大利），David Cuartielles（西班牙），David Mellis、Tom Igoe（美国）。Arduino 是一个开放的硬件平台，包括一个简单易用的 I/O 电路板，以及一个基于 Eclipse 的软件开发环境。Arduino 可以用来开发可独立运作并具互动性的电子用品，或者也可以开发出与 PC 相连的周边装置，同时能在运行时与 PC 上的软件（如 Flash、Max/Msp、Director、Processing 等）进行通信。

Arduino 之所以会得到创客们的广泛青睐，这主要归功于它的两大特性：易用性与开源性。

相比于其他微控制器开发平台，Arduino 最大的优点是它的易用性。首先，入门门槛低。我们不需要任何硬件电子知识，也无须了解硬件内部结构以及寄存器设置。同样，我们也不需要熟悉复杂的底层代码以及晦涩的汇编语言编程知识。只要我们知道 Arduino 的端口功能及其调用函数便可以进行编程开发，便可设计出令人惊艳的互动装置！其次，扩展性强大。有相当多的和 Arduino 兼容的扩展板可以直接插在 Arduino 开发板上使用，几乎包括了你能想到的每个领域，这样可以避免使用锡焊，而是简单地把扩展板一个叠一个地插在一起。

Arduino 能够获得广泛认可的另一个关键因素是，它的开发板设计以及程序开发接口都是开源的。也就是说，Arduino 所有的设计都是可以免费获得的。如果你愿意，可以直接下载开发板电路设计图，从商店购买所需的电子元件自己完成 Arduino 开发板制作。开源的 IDE（集成开发环境）可以免费从官网下载。此外，互联网上有许多活跃的 Arduino 论坛，上面有大量其他人开发的开源作品，既可以作为自己的参考也可以直接拿来使用，而且随时可以在论坛找人帮助。

总之，凭借其简单的接口调用以及强大的功能，Arduino 得到了众多电子设计爱好者的追捧。而其廉价、开源的软硬件支持，更是极大地促进了电子互动设计的发展，为智力分享创造出广阔的沃土，从而让越来越多的人高效地设计出惊艳、实用的作品。

7.4.2 初窥 Arduino

7.4.1 节我们简要了解了 Arduino 的地位、功能以及优势，然而未对其本身做出介绍，那么 Arduino 到底是什么呢？我们先看一下维基百科的描述：“Arduino 是一块单板的微控制器和一整套的开发软件，它的硬件包含一个以 Atmel AVR 单片机为核心的开发板和其他各种 I/O 板，软件包括一个标准编程语言开发环境和在开发板上运行的烧录程序。”

简单地说，Arduino 包含两个主要的部分：硬件部分是用作电路连接的 Arduino 开发板（如图 7-20 所示），软件部分则是用来编写控制程序的 Arduino 集成开发环境（如图 7-21 所示）。

Arduino 开发板是由一个小型微处理器和一个电路板所构成的。比如，我们可以采用 ATmega328 来作为整块电路板的核心，它是一枚具有 28 个引脚的黑色细长芯片，它就像一个微型计算机，其运算能力虽然无法与 PC 机相比，但是完全可以胜任一般的互动装置开发，并且价格十分便宜。电路板上已经焊接了该芯片正常工作所需的所有电子元件，使用 USB 线将开发板连接到计算机上后，随时可以进行通信。此外，不同的开发板代号各不相同，本书示范时使用的是当前 Arduino 最新版本开发板，代号为 Arduino UNO R3，如图 7-20 所示。

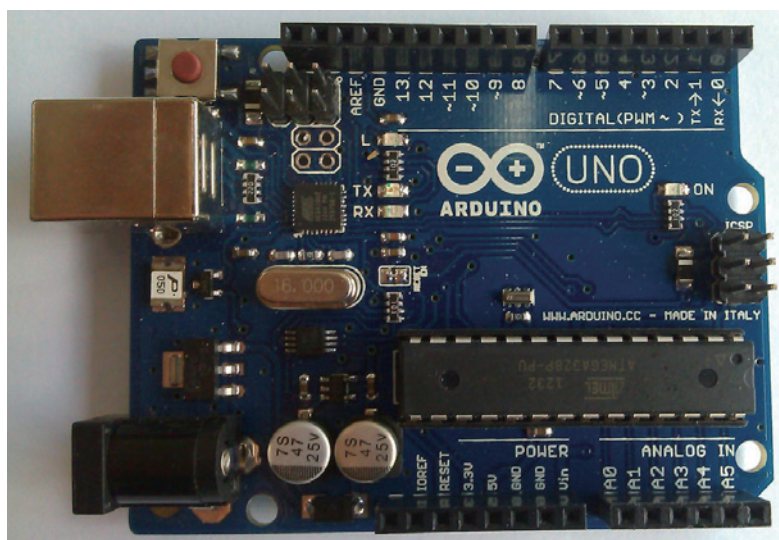


图 7-20 Arduino UNO R3 开发板

在图片中，我们可以看到 Arduino 开发板上有很多引脚，你可能会感到困惑。下面就来解释每个部分的作用和功能。

- 标有 0 ~ 13 的一排共 14 个引脚，均被称作数字 I/O 引脚，它们可以在程序中被自由设定用于输出或输入数字信号。此外，3、5、6、9、10、11 号引脚也被称作模拟输出引脚，因为它们均可由程序指定从而变更为模拟的输出引脚。
- 标有 A0 ~ A5 的一排共 6 个引脚，均被称作模拟输入引脚，这些引脚用于读取各种模拟输入信号（例如，读取超声波测距仪的脉冲响应时间），并在程序中将其转化为 0 ~ 1 023 的数值。
- 紧邻模拟输入引脚的另一排共 6 个引脚称作电源接口，第一位是重置（Reset），连接低电平可重启单片机，使程序从头开始运行。其他的接口用于提供不同的电压（3.5V、5V、GND 和 9V），其中，GND 又称地线，是开发板上其他电压值的参考值。

Arduino 集成开发环境（IDE）是一个在计算机里运行的特殊软件，如图 7-21 所示，我们通过它来给 Arduino 开发板上传不同的程序。而在使用 Arduino 进行开发前，必须先下载集成开发环境，地址是 www.arduino.cc/en/main/software，然后按照你的计算机操作系统选择下载版本，相应的安装指南可以在 Arduino 官网（www.arduino.cc）上找到。

例如，在 Windows 7 系统下，如果下载了最新的 IDE，解压文件并双击打开已解压文件夹，你将会看到 Arduino 文件和子文件夹在里面。然后，用 USB 电缆连接计算机与 Arduino 开发板（此时标志为 PWR 的绿色电源 LED 灯亮），接着 Windows 将会为 Arduino 开发板自动安装驱动程序。

至此，我们已经对 Arduino 有了一个基本的了解，并且开发前准备工作基本就绪，接下来我们就可以开始发挥自己的才智，设计我们自己的 Arduino 互动装置。

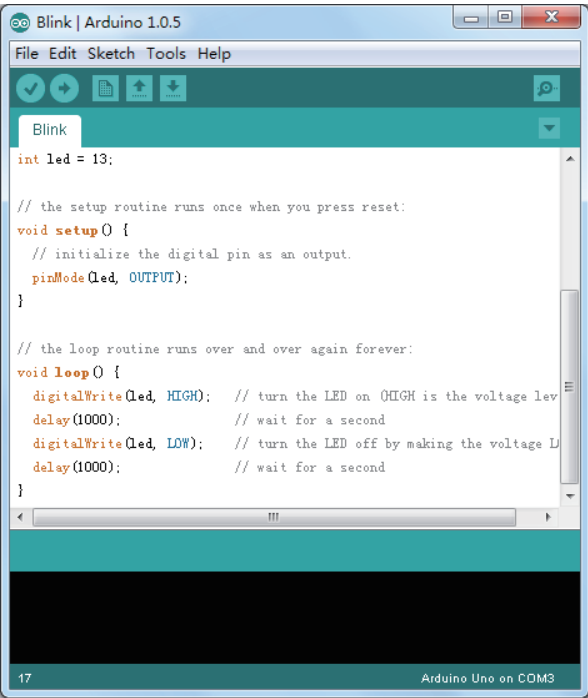


图 7-21 Arduino 集成开发环境（IDE）

7.4.3 牛刀小试：叩开 Arduino 之门

LED 闪烁程序是 Arduino 初学者的第一个程序,如同程序设计语言中打印“Hello,World”一般。此外,我们也可以用它来测试 Arduino 开发板是否能够正常工作。

所需元件:面包板、5mm LED、100 Ω 电阻、导线若干。当然,如果有传感器扩展板(如图 7-22 所示)以及 LED 集成模块(如图 7-23 所示)则无须这些分离元件。

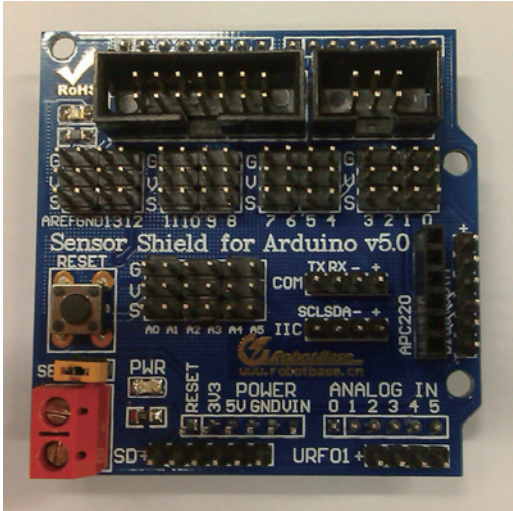


图 7-22 传感器扩展板

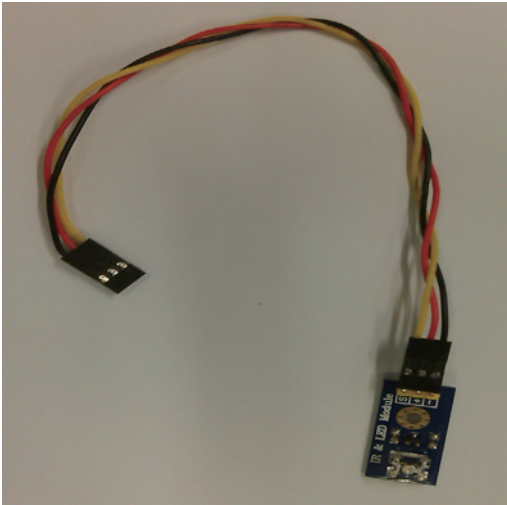





图 7-23 LED 集成模块

电路连接:取出 Arduino 开发板,选择 13 号数字 I/O 引脚(0 ~ 13 号数字信号引脚可任意选取) 作为数字信号输出引脚,依次将导线、电阻、LED、GND 引脚串联起来形成一个回路。最后,通过 USB 电缆连接 Arduino 与计算机。

环境设置:首先,设置“Tools”中的“Board”型号。比如,若使用如图 7-20 所示的开发板,则必须选择“Board”中的“Arduino Uno”。其次,设置“Tools”中的串行口“Serial Port”。查看串行端口号的方法:依次单击 Windows 的“开始”→“控制面板”→“系统”→“设备管理器”→“端口 (COM 和 LPT)”,便可查看系统分配给 Arduino 的端口号。

代码编写:按照 Arduino IDE 安装路径,双击打开 arduino.exe,并将图 7-21 中的代码输入 Arduino IDE 中。或者单击按钮选择“01.Basics”中的“Blink”,便可打开 Arduino IDE 中自带的该 LED 闪烁代码。

代码上传:首先单击按钮, Arduino 软件窗口底部出现“Done Compling”,接着单击按钮, Arduino 软件窗口底部出现“Done Uploading”,即表示程序上传成功:程序代码已经上传到 Arduino 开发板的 ATmega328 芯片上,等待几秒,Arduino 开发板就会完成自动更新。

实验结果:程序上传的过程中,Arduino 开发板上的 LED 会一阵狂闪,几秒以后外接的 LED 以 1s 的间隔闪烁,如图 7-24 所示。

此时,我们已经叩开了 Arduino 的大门,完成了探索、利用 Arduino 的第一步。虽然上面这个实验非常简单,但是基本展示了 Arduino 开发的所有流程,总结如下。

1. 按照设计功能需求,准备好所需元件。
2. 将 Arduino 开发板、相应扩展板以及其他元件电路连接起来。
3. 通过 USB 线将 Arduino 与计算机相连。
4. 开始编写代码并调试。
5. 将完整、正确的代码上传到 Arduino 开发板上。
6. 等待几秒,Arduino 会自动更新并开始工作。

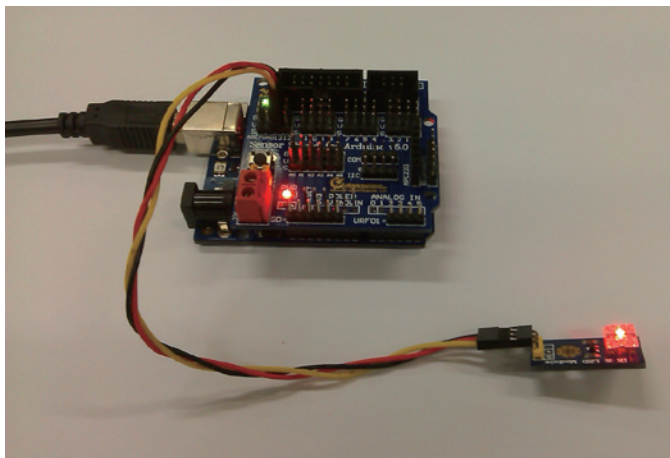


图 7-24 Arduino 控制 LED 闪烁实验

7.5 创客之开源软件Android（安卓）

根据 7.4 节的学习，我们知道 Arduino 开源硬件能够与许多功能强大的扩展板以及大量的电子元器件相连接，并且对它们进行控制以完成一系列强大的功能。但是，它有一个缺点，即缺乏一个良好的人机交互界面，并且无线连接能力也很差。然而，值得庆幸的是，Android（即安卓，你的三星、小米、华为、联想手机上用的都是它）开源软件系统能够有力地弥补这个缺陷：它不仅有一个良好的人机交互界面（一个超大的触摸屏），而且是开源的，允许我们设计出完全满足自己需求的个性化的功能。本节我们将一起来了解 Android 以及它的开发过程，并给出一个利用开源程序完成的开发实例。

7.5.1 Android 概述

Android 是一种基于 Linux 内核的、自由的并开放源代码的操作系统，如图 7-25 所示，主要用于移动设备，如智能手机和平板电脑。Android 操作系统最初是由 Andy Rubin 开发，主要支持手机。2005 年 8 月由 Google 收购注资，得到了进一步完善。2007 年，Android 迎来了一个里程碑式的发展，Google 与 84 家硬件制造商、软件开发商及电信运营商组建开放手机联盟，共同研发并改良 Android 系统，随后便发布了 Android 源代码。



图 7-25 Android Logo

归纳起来，Android 具有如下一些突出优势。

- 开放性。这是 Android 最鲜明的一个特色，开放的平台允许任何移动终端厂商加入到 Android 联盟中来，从而为 Android 创造一个广阔的市场。而显著的开放性还可以使其拥有更多的开发者，这样 Android 可以吸收无数程序员的思想精华，不断壮大、快速走向成熟。
- 自由性。Android 开发平台给第三方开发商提供了一个十分宽泛、自由的环境，不会受到各种条条框框的阻挠，不仅可以利用免费 Android 平台进行开发，还可以在该平台上推广自己的服务。可想而知，在这个平台上会有大量新颖别致或极具个性化的软件。
- 强大的功能。由于 Android 系统是以 Linux 内核为基础的，所以它继承了 Linux 最显著的功能优势：高效性与灵活性，并且 Android 通过引入中介层方式使得它在移动设备上获得了更高的效率；具有多任务能力，多个软件可以同时并独立地运行；同时具有字符界面与图形界面，用户体验良好；支持多种硬件平台，即可以运行在手机、平板电脑、机顶盒以及游戏机等众多设备上。
- 丰富的服务。诞生于互联网的 Google 已经走过了 10 年的历史，从搜索巨人到全面的互

联网渗透，Google 服务如地图、邮件、搜索等已经成为连接用户和互联网的重要纽带，而 Android 平台设备将无缝结合这些优秀的 Google 服务。比如，在 Android 手机中用户只需登入自己的 Gmail 地址，即可直接通过 Android 系统的内置程序使用这些服务。

正是因为 Android 特有的巨大优势，它在全球智能手机操作系统市场上的份额迅速扩张，并成为当前最主流的手机系统。根据市场研究公司 Canalsys 的数据显示，2009 年第 2 季度 Android 占据全球智能手机操作系统市场 2.8% 的份额，而在 2010 年第 4 季度的全球份额中增长到了 33%，Android 操作系统也因此击败了诺基亚的 Symbian 系统、苹果 iOS 系统成为了全球第一大智能手机操作系统。2012 年 5 月，根据市场调查公司的数据显示，Android 操作系统在全球智能手机操作系统中的份额已经过半，达到了惊人的 60%。2012 年 6 月，Google 在 2012 Google I/O 大会上表示全球市场上有 4 亿部 Android 设备被启动，每天启动一百万台。2013 年 5 月，Android 在中国的占有率有 71.5%，超过其主要竞争对手苹果公司约 50%，而世界占有率亦有近 70%。

总之，Android 的开源性以及功能上的优越性，不仅吸引了众多大厂商的支持与推广，而且获得了全球广大软件开发人员的青睐，从而使 Android 得以迅猛发展并成为当前最主流的手机操作系统。毫无疑问，Android 将继续引领手机操作系统，这也必将为无数开发爱好者提供越来越大的创作舞台，推动技术的不断进步。

7.5.2 开发平台搭建

从 7.5.1 节的介绍当中，我们得知 Android 是一个非常强大的操作系统。正因如此，目前已经出现了大量运行于 Android 手机系统之上的应用软件，如图 7-26 所示。有数据显示，截至 2011 年 10 月，Android 市场上已有超过 30 万个应用程序，并且在 2011 年 12 月，Android 市场上的应用程序下载量超过 100 亿次。



图 7-26 (a) Android 手机上安装的应用程序, (b) Android 市场应用软件

Android 市场有大量的应用软件可供下载，其中，有的方便了我们的生活，如公交线路查询、手机支付应用等；有的为我们提供休闲娱乐，如一些视频、歌曲以及社交网络应用等；有的则能够帮助我们学习，如一些读书、考试模拟等应用；还有一些更有用、有趣的应用，比如控制窗帘、家电等。然而，在享受娱乐、便利之余，我们不禁会想，这些设计精美、功能强大的应用软件怎么设计出来的呢？我们自己能不能也设计一些有创意的、满足我们个人需要的应用软件呢？答案是肯定的。不过，要想开发出 Android 应用程序，我们就必须要有完备的开发工具。所以，下面我们就开始着手准备开发工具，搭建开发平台。

Android 开发平台在 Windows、Linux、Mac 操作系统中均可完成搭建，但方法有所不同，我们以主流操作系统 Windows 为例介绍 Android 开发平台搭建过程。

Android 开发平台是由多个开发包组成的，具体说明如下。

- JDK：即 Java SDK。Android 开发是以 Java 作为开发语言的，而所有用 Java 开发的应用程序都需要安装 Java 虚拟机。下载地址：<http://java.sun.com>。
- Eclipse：Eclipse 是一款免费、优秀的开源集成开发环境（IDE），很多 Java 项目开发都基于该平台，当然 Android 应用程序开发也不例外。下载地址：<http://www.eclipse.org>。
- Android SDK：即 Android 软件开发工具包，是应用软件开发工具的集合。该工具包定义了许多 Android 手机开发的底层应用，可以调用这些底层工具实现更多、更复杂的手机应用。下载地址：<http://developer.android.com/sdk/index.html>。
- ADT：将 Eclipse 和 Android SDK 连接起来的纽带，在 Eclipse 编译 IDE 环境中安装 ADT，为 Android 开发提供开发工具的升级或变更。下载地址：<https://dl-ssl.google.com/android/eclipse/>（可在线安装，不必下载）。

当上面的 4 个工具包已经下载并安装就绪后，我们就可以开始 Android 开发平台的正式搭建了，步骤如下。

第 1 步，安装 Java SDK 并配置 Java 开发环境。

第 2 步，Eclipse 开发工具安装。

第 3 步，Android SDK 的安装与配置。

第 4 步，ADT 安装与配置。

第 5 步，在 Eclipse 中设定 Android SDK 目录。

7.5.3 Android 之旅起航：Hello, Android!

在早期的 Android 应用程序开发中，通常通过在 Android SDK（Android 软件开发包）中使用 Java 作为编程语言来开发应用程序。开发者也可通过在 Android NDK（Android Native 开发包）中使用 C 语言或者 C++ 语言作为编程语言来开发应用程序。在此，我们以 Java 为开发语言，用一个最简单的开发实例来介绍 Android 开发流程。

1. 创建模拟器。因为我们在开发阶段并不需要用到真正的手机，而是先在计算机上模拟程序的功能，即在“模拟器”中进行仿真。

(1) 打开 Eclipse。选择 “Windows” → “AVD Manager”。这时会弹出一个窗口，单击右侧 “New” 按钮，弹出如图 7-27 所示的对话框。

(2) 单击 “Create AVD” 按钮完成创建模拟器。

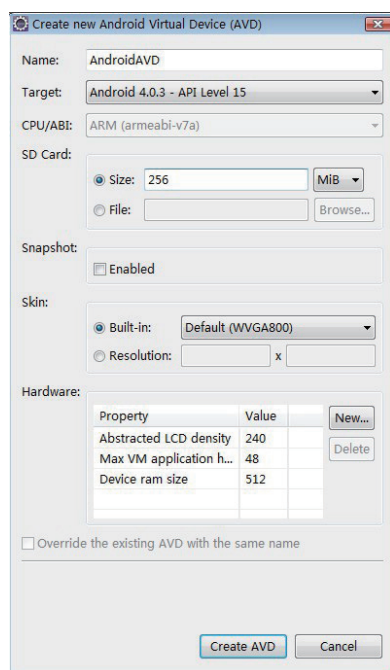


图 7-27 新建一个模拟器

2. 开启模拟器

(1) 选中刚刚建立的模拟器，单击右侧的 “Start...” 按钮来启动模拟器（如图 7-28 所示）。

(2) 经过几秒钟等待，模拟器开启，弹出如图 7-29 所示的界面。

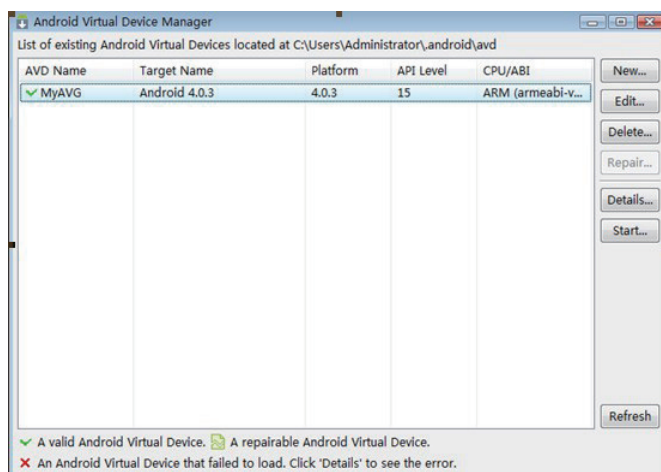


图 7-28 选择并启动新建的模拟器

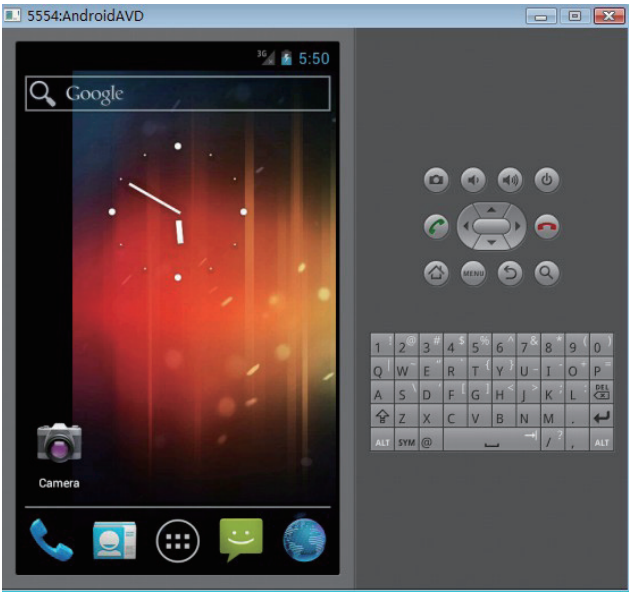


图 7-29 模拟器界面

3. 创建名为“HelloAndroid”的程序

(1) 模拟器开启之后，下面我们就可以创建第一个 Android 程序了。

(2) 选择“File”→“New”→“Project”命令，建立新项目“HelloAndroid”。

(3) 填写相应的参数，如图 7-30 所示。需要注意的是，必须输入一个“Package Name”（包名），否则创建不了 Android 程序。

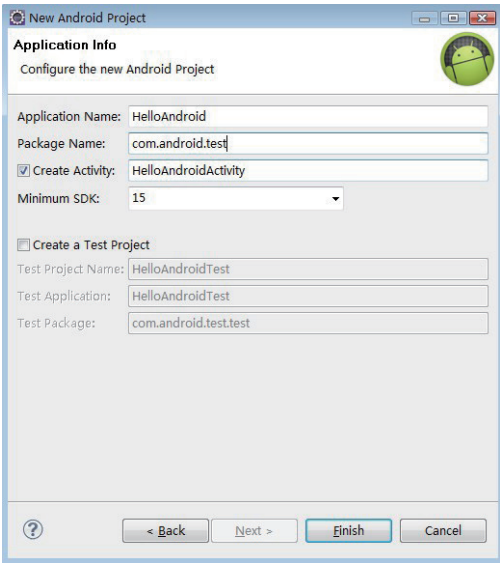


图 7-30 创建新项目“HelloAndroid”

创建完成之后，主界面左侧解决方案栏出现如图 7-31 所示的列表。

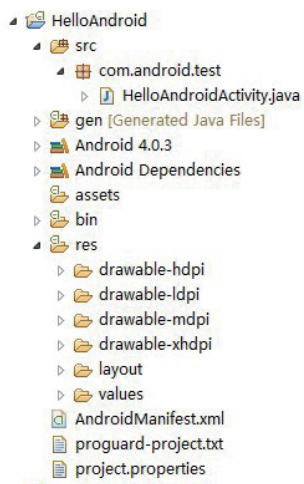


图 7-31 解决方案栏显示列表

4. 修改“res”文件夹下“values”子文件夹中的“strings.xml”文件，对标签“hello”的值进行修改，如改为“Hello Android, I am coming!”，如图 7-32 所示。

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <resources>
3 <string name="hello">Hello Android,I am coming! </string>
4 <string name="app_name">HelloAndroid</string>
5 </resources>
```

图 7-32 修改“name”值为“hello”的标签值

5. 选中“HelloAndroid”项目，鼠标右击“Run AS”→“Android Application”运行程序，效果如图 7-33 所示，手机界面中显示着刚才那句“Hello Android, I am coming!”。至此，一个简单但完整的 Android 应用软件开发就基本完成了。是不是确实非常容易上手？

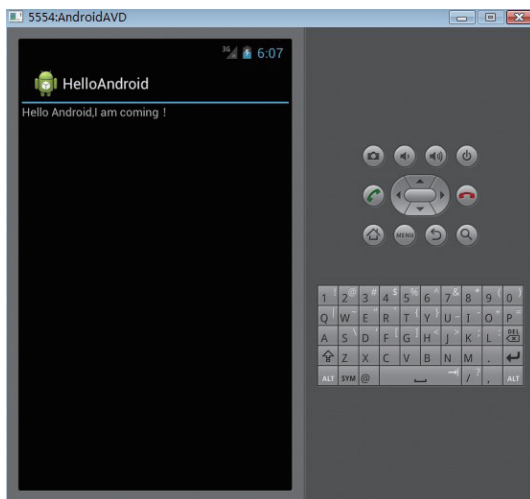


图 7-33 HelloAndroid 运行效果图

7.6 靠创意去赚钱：漫谈Kickstarter、Quirky与Shapeways

创客的核心在于创意，然后才是DIY出来，把创意变成现实。当然，很多创客并不满足于做出产品，还会想着进一步获得赢利，以便补贴家用乃至以此创业。下面，我们就重点介绍国外如火如荼的3个网站Kickstarter、Quirky、Shapeways，分别代表了将创意转化成Money（真金白银）的3种典型模式，以供创客们参考。

7.6.1 Kickstarter 众筹：靠创意去筹资

Kickstarter 是一家以**众筹（Crowdfunding）**的方式为创造性项目筹集资金的网站平台。众筹，即大众筹资或群众筹资，是指用团购+预购的形式，向网友募集项目资金的模式。众筹利用互联网和SNS传播的特性，让小企业、艺术家或个人对公众展示他们的创意，争取大家的关注和支持，进而获得所需要的资金援助。

Kickstarter 于2009年4月在美国纽约创立，该网站的诞生与其中一位华裔创始人Perry Chen（佩里陈）的经历不无关系。他的正式职业是期货交易员，但因为热爱艺术，开办了一家画廊，还时常参与主办一些音乐会。2002年，他因为资金问题被迫取消了一场筹划中的在新奥尔良爵士音乐节上举办的音乐会，这让他非常失落，进而就开始酝酿建立起一个募集资金的网站。

Kickstarter 虽然不是最早以众筹概念出现的网站，但却是最先做成的一家，曾被《时代周刊》评为最佳发明和最佳网站，进而成为“众筹”模式的代名词。在Kickstarter网站上，每个人都有可能让自己的梦想变成现实。目前，Kickstarter 已经是全球最大的创意项目募资平台。截至2013年5月，Kickstarter 上共发布有10万个项目，其中43.95%筹钱成功，共计5亿3500万美元。单个项目筹集到的最高资金额为1000万美元。

Kickstarter 平台的运作方式相对来说较为简单而有效：该平台的用户一方是有新创意、渴望进行创作和创造的人，另一方则是愿意出钱、帮助他们实现创造性想法的人，然后见证新发明、新创作、新产品的出现。Kickstarter 提供了“有创意、有想法，但缺乏资金”与“有资金，也愿意捐款支持好创意”的平台。

Kickstarter 让商品和服务提供者变成了有梦想要实现的人，让消费者变成了愿意资助梦想的好人。一些新奇的创意和设计在这个网站上变成了现实，在这个过程中，Kickstarter 进一步强化自己是个梦想平台的概念，甚至被称为“一个神奇的网站”。Kickstarter 上的项目五花八门，有漫画书出版计划、音乐、盲文手表、纯手工糕点，还有令人扼腕称奇的3D打印机。你在Kickstarter上可以看到那些在亚马逊和新蛋的网站上永远不会出现的奇特产品：包括自动掌握平衡的双轮摩托车、专门复制家具的3D打印机、彩色的异型计算机主机，以及各种小众电影和游戏等。

Kickstarter 致力于让你忘掉购物这码事，在这个众筹网站上，你花出去的钱不叫购物，而被称为“资助了一个梦想”，你得到的不是单纯的服务或商品，而是一个故事发生的过程。用户因为了解蛋糕制作人的故事，连带也觉得这块纯手工蛋糕与众不同。

对于创业者来说，Kickstarter 不仅意味着钱。大多数Kickstarter的项目都只是一个设计稿，或者是一个不完善的产品雏形。而一些项目的出资人等于预购了这款产品，成为忠实的“天使

用户”。认购的用户可以在项目完成之后以较低廉的价格优先获得产品，但同时也承担了项目启动却最终未能形成产品的风险。

产品开发团队需要向出资人定期发布产品开发进度，并对相关建议和意见做出回应。对于一个全新的产品来说，这是一次难得的市场前期调研。同时，开发团队也可将遇到的问题提交到网站上来获取意见和帮助，提供帮助和解决方案的人来源于项目相关领域的爱好者或者是专业的技术人员，而这类人力资源的使用是无偿免费的。

Kickstarter 和传统的投资途径完全不同。项目创造者可以选择一个融资截止时间和一个最低融资目标金额。比如，加州马金·卡拉汉希望创作一部关于半人半妖的新漫画，第一期的创作和宣传费用预计需要 1 500 美元，因此，她在网站上开启了一个项目，希望有人能够提供小额捐款。捐款者可以得到的回报是，捐 5 美元可以得到一册带有作者签名的漫画书，捐 100 美元可以得到一个带有以漫画故事中主人公为饰物的包。当然，只有收到的总捐款超过了目标金额 1 500 美元，她的许诺才会兑现。结果是，她在很短的时间里就拥有了这笔捐款。

Kickstarter 采用一种“质押转辙”机制，在一定程度上保证了投资人的资金安全。投资人承诺所投的资金将通过第三方支付机构——亚马逊支付（Amazon Payment）筹集在一起，该平台对全世界所有对众筹项目感兴趣的投资人开放。Kickstarter 会在每笔成功的众筹融资中抽取 5% 的佣金，亚马逊将抽取另外 3% ~ 5% 的佣金。和其他众筹融资平台不同，Kickstarter 表示他们不会拥有任何创意项目的所有权。所有在该网站进行众筹交易的项目网页将会被永久性存档，方便大众访问。一旦众筹项目完成，项目内容和上传的媒体信息将无法再次编辑，也无法删除。Kickstarter 采取一刀切的方式，任何无法在有效时间内筹集到目标资金的项目必须把钱退还给支持者。也即，如果一个众筹项目在融资截止时间到达时没有达到最低融资目标，该项目的融资即告失败，已筹集的资金也无法获得。

目前，我国还没有一家众筹网站能像 Kickstarter 那样成功，有些甚至只能说是勉强存活着。相比于 Kickstarter 时常爆出一些超级热门的集资项目，国内的众筹平台到目前为止几乎还没有一个真正拿得出手的项目，我们也没听到有哪个国内的众筹项目受到大众的追捧并获得超额募资。

众筹网站目前之所以还没有在我国国内火起来，原因不外乎如下几点。

- 募捐这种方式在国内还没有被广为接受，而在欧美，募捐是很常见的，大到选总统，小到很小的活动。
- 几例非法集资案的判决让人对募捐这种方式心存畏惧。
- 知识产权的保护不成熟，创意难以转化为收益。
- 创新力不足，没有真正让人震撼的创意项目。

但是，不得不说，众筹这种模式的确是让人耳目一新的，期待在国内会有更好的发展。当然，在政府和国家支持方面，需要配合有相关的保障政策和法规出台。

7.6.2 Quirky 创意加工厂：把创意变成产品

Kickstarter 称“对网站上的项目最终决定权在用户手里”，言外之意也就是你要对自己的

投资决定负责。另一家众筹模式的网站 Quirky 直接瞄准硬件和设计产品这个细分领域。它跟 Kickstarter 的不同之处还在于，它努力维护着一个有专家氛围的创意挑选社区，并组建了一支专业的团队负责最终决定和生产阶段的执行。从产品生产和销售环节看，Quirky 俨然就是一家创意设计产品的电子商务（简称电商）公司，只是通过把设计阶段众包出去，使得它能提供与其他 B2C 平台上不同的新奇产品，这也是它与其他设计品电商网站相比的竞争力所在。



提示：众包（Crowdsourcing）指的是一个公司或机构把过去由专业员工执行的工作任务，以自由自愿的形式**外包（Outsourcing）**给大众网民的做法。与强调高度专业化的传统外包不同，众包通过跨专业的大众网民的集体智慧来获得超乎寻常且低成本的创新。典型的案例是维基百科和百度百科的诞生，由众多个人用户自愿创作的词条已达《大英百科全书》的几十倍，甚至上百倍。

与 Kickstarter 只为你实现想法募集资金的模式不同，Quirky 更进了一步，直接帮你把想法实现成可用的产品。

在 Quirky 上活跃的大多数是一些业余的开发爱好者、学生、退休人员以及热衷产品设计的个体。用户提交的创意点子通常由其他用户投票，每月该网站都会选出得票最多的两个创意，供以工程师及设计师等组成的一个内部团队进行研究，由他们做出决断，并将有价值的点子开发出雏形。在构思、设计、命名、打造 Logo、包装等各个环节中，这个社区都扮演着协调者的角色，让社区用户各抒己见，协作参与整个开发全程。

Quirky 的工作模式是这样的，比如说，我有一个想法是要做一台时光机器（以便穿越到明朝的北京去买一个二环边上的小四合院）。我把想法提交给 Quirky（支付 10 美元），Quirky 社区的人很快就会给我提出意见和反馈，这样我的想法就可以得到改进，并能够实现。接着，Quirky 小组会对那些最有前途并能商业化的产品做进一步的研究、设计和品牌构想。

接下来就更有意思了。Quirky 的产品设计师和工程师组成的专家组会将这个改良后的创意做成一个原型产品。再经过各方面精细的调整，小批次的产品就做出来了！他们会把这批产品拿来预售以做测试。如果预售表现不错，他们就确定了这一产品的市场需求，接下来就是量产了。

这么一来就更正规了！Quirky 将会努力扩大其销售渠道（社会渠道、直销以及零售渠道）。那么我们能收获什么呢？发明人（这个例子里的我）、提供帮助的社区，当然还有 Quirky，都会分享产品销售的利润，这真是太棒了！Quirky 是第一家为参与的开发者提供现金奖励的公司。那些对产品设计、生产以及销售有所贡献的人，都将获得现金奖励。参与的形式包括投票、评论、打分、搜索、在预售期进行评测等。社区将会获得产品线上直销利润的 30% 以及线下销售利润的 10%，然后再根据个人的贡献进行分配。不少活跃分子已经获得了上万美元的收益。这种商业模式直接铸就了 Quirky 的成功。

Quirky 的一个典型成功案例是一位中学生创客杰克·兹恩（Jake Zien）所设计的可自由改变形状的链式插座 Pivot Power，如图 7-34 所示。兹恩在 Quirky 平台提交这个点子仅仅花费了 10 美元（提交费的设置是为了杜绝有些用户以无聊的态度来胡乱张贴内容，比如笔者前面说的时光机器），然后便进入与 Quirky 内部小组及全体用户的开发讨论阶段。最后该产品在线上售卖，同时向 Bed Bath & Beyond 连锁商店及电视购物公司 HSN 等渠道铺货。在产品上线的第一周，兹

恩便得到了 2.8 万美元的回报，这是一个十分令人满意的数据。2012 年该产品共获得了 50 万美元的净收入，而兹恩本人的收入则超过 5 万美元。



图 7-34 Quirky 的一个典型案例：可自由改变形状的链式插座 Pivot Power

据统计，将一个创意推进至成品的前期过程中，一般公司需要耗费 20 万美元的研发费用，这还只是按图纸生产出第一个产品的费用。正是因为这中间有巨大的投资风险，所以大多数人不敢将发明设想化为实践。将创意带入研发程序，助力其迈开第一步，是 Quirky 的伟大之处。

7.6.3 Shapeways 在线打印：把个性化产品定制出来

Shapeways 是一家创新制造公司，利用 3D 打印技术为客户定制各种产品，包括艺术品、精巧又复杂的珠宝挂饰、iPhone 手机壳、玩具、杯子，还为设计师提供了销售其创意产品的网络平台。这家公司已获得了数千万美元的风险投资。

这家提供在线打印服务的公司，并不直接出售打印机，而是打造了一个 3D 打印服务的社交网络。用户在网站注册后，既可以把自已的产品设计上传到网站，也可以购买现有的 3D 设计图，再选择和购买原材料，之后就可以下单，然后 Shapeways 利用设在纽约“未来工厂”里的 50 台工业级 3D 打印机将其打印出来并邮寄给用户。此外，设计师也可以在线展示和销售产品，并将产品卖给其他人。

从根本上说，Shapeways 是即时生产，流程是客户下单后再快速生产，然后尽快寄给客户。Shapeways 赋予了产品个性化的特点，比如客户可以把自己喜欢的一些要素融入产品，或是完全从头按自己的想法设计产品。Shapeways 相当于利用自己的专业设备给普通用户代加工。虽然目前的家用 3D 打印机已很便宜，但维护起来并不轻松，更不要说那些造价昂贵且维护复杂的专业设备了。Shapeways 为用户提供了 30 多种专业材料选择，包括常见的玻璃、金属（包括金、银），还有一种弹性良好的塑料。此外，还可以打印各种陶瓷产品，你可以利用这项服务打印出自己的陶瓷餐具和茶具。

Shapeways 除了能为你打印 3D 设计之外，它也提供了一个市场，让设计师可以销售定制的 3D 打印产品给客户。事实上，Shapeways 的运作流程相当简单：Shapeways 提供一个该产品的生产成本价，在此基础上设计师根据自己的意愿设定一个销售价，Shapeways 直接把设计师定价的产品寄给买家，并在月底把货款打给设计师。

有了 Shapeways，事情就变得很简单了。创业者有了创意之后，只需进行设计和测试，就能在 Shapeways 上免费开店、让广大消费者看到自己的产品。无论你想生产 10 个还是 1 000 个，Shapeways 都不介意。因此越来越多的小公司利用 Shapeways 的服务，在网站上开设在线商店。自 2007 年成立以来，Shapeways 已经生产了超过 100 万款 3D 打印产品，每月新增 6 万多款，总产量已超过 60 亿件，在线商店的数量多达 8 000 家以上。

传统的制造工厂通常对订单有最低数量要求（例如 1 万件起），因为模具的成本非常昂贵。而由于 3D 打印无须模具的特点优势，Shapeways 甚至对只做 10 个的小订单也非常欢迎。Shapeways 网站的收入来源为设计师（个人厂长）缴纳的手续费。设计师能够自行决定他们在这个网站上销售的产品价格，但是其价格中扣除原材料费用之后的利润的 3.5% 将支付给 Shapeways 网站。

7.7 创客中国：中国版乔布斯和比尔·盖茨的诞生地

“创客”泛指那些热爱动手实践，努力把各种创意转变为现实的人。创客并非今天才出现，乔布斯和他的搭档沃兹尼亚克，还有比尔·盖茨和他的搭档保罗·艾伦都是典型的创客。随着软件、硬件、制造、艺术的日趋结合，创客们有了更大的活动舞台，不再像上一代的黑客那样局限于软件编程和网站开发，而能在比特世界、电子世界、原子世界、艺术世界四者之间自由穿梭跨界，打造出一个个令人炫目的创新产品。

尽管创客们最开始都以好玩为主要目的，但当创意及其实现有成为商业模式的可能时，创业就是一件顺理成章的事情。一旦有创业的想法，就要去思考商业模式，组建创业团队。从创意到实现创意是一个质的飞跃，从创意产品到形成商业模式，又是一个飞跃。未来几年，创客人群与主流商业的融合将会成为必然。

7.7.1 国外创客为什么纷纷青睐中国

目前国际上有五家最大的开源公司，其中两家就在中国，一家在上海，一家在深圳。这两家公司，在 3 年前都进入了开源硬件的领域，所做的产品都可以卖很高的单价，这是因为他们有自己的高技术创新，并形成了品牌，拥有国际化的粉丝群体。

MakerBot 公司的创始人 Zach Smith（扎克·史密斯）来中国深圳已经有几年了，什么原因让扎克来深圳创业？2011 年 4 月，扎克和 4 位同事来到广东，筹备在中国生产 MakerBot 3D 打印机。遗憾的是由于 MakerBot 管理层内部冲突，扎克被迫离开。不过事后扎克并未返回美国，而是决定加入创业孵化器 HAXLR8R，留在深圳帮助专注硬件产品的海外创业团队打磨产品。他觉得这里才是最适合工程师、创客和硬件创业者的地方，甚至超过硅谷和纽约。而扎克来深圳创业，看中的是深圳的效率和成本优势。“深圳不仅有高密度的供应链和生产，蕴藏的生产经验也同样丰富。”这是扎克对深圳的评价。

国外创客们纷纷来到中国落脚的逻辑很简单：创客的本质是把想法做成产品，这些想法成为产品后并非百分之百靠谱，所以需要验证。3D 打印机的意义正在于此，能够在设计环节快

速打印出模具，但是一旦涉及做出产品或者进行小批量生产，硅谷的创客们会发现他们所需要的各种原件几乎都来自中国，索性干脆常驻中国，因为在这里他们可以仅用一天的时间在一个电子市场内以极低的价格买到所有他们需要的原件，如果产品有小规模量产的需求，这里还有成熟的产业链条可以对接。一些硅谷的冒险家也来到深圳充当掮客角色：一旦某个项目上了Kickstarter并募资达到可量产规模，他们就会发一封邮件过去，询问是否需要帮忙联系在中国的工厂来做产品的代工生产。

7.7.2 创客中国的背景优势

在德国、日本等制造强国，生产资源通常只会留给每批次达成千上万规模的订单。从研发到生产，创客作品的工业化在国外遭遇巨大壁垒，而只有中国才具有足够分散的生产能力，愿意承接创客数百量级甚至更小的订单。因此，中国新时代的技术创新者，不再需要像互联网时代 Copy to China（将国外的成功模式复制到中国），生产制造优势能让我们以更快的速度做研发，更低的成本做制造，而且中国有非常庞大的市场，拥有整合生产链的能力。以之前的“山寨”为例，其实质是未经授权的微创新，它可以快速生产、快速调整。而传统大公司进行硬件研发要一到两年，这已经不适合互联网时代的用户需求了。当然，随着创客以及中国智造的产业升级，山寨文化终将成为翻过去的一页，创客中国将取而代之。

在国内，创客最集中的地方当属北京、上海、深圳这三座城市。以深圳为例，它是产业链最完善的城市，一个创客来到这里可以完成从产品原型到做出产品再到小批量生产的整个过程，这里既有创业氛围浓厚的柴火创客空间，又有 HAXLR8R 这样的硬件加速孵化器，也有 Seeed Studio 这样的能够为创客提供小批量生产的组织。此外还有著名的华强北，在这里采购硬件原料的成本非常低，并且可以随时找到所需的材料，卖给世界“任何一端”的客户。

国内的一家初创公司将他们设计的机器人套件 Makeblock 搬上了美国的众筹网站平台 Kickstarter，项目截止时融资额达 18 万美元，是预期的 6 倍，远远超出了创始人王建军的预料。王建军的团队之后就充分利用了中国制造业柔性化供应链和成本低的优势，让“创新硬件+批量生产”这一模式变得简单很多。

“国内的工厂工艺不复杂，而且愿意接小订单，能快速跟上研发进程，这些都是中国制造业的优势。但在德国等海外市场，制造业自动化程度高，而且不做小批量生产。”王建军说。确实，深圳拥有成熟的电子制造产业，周边城市如东莞、佛山又具备强大的机械加工能力，弹性化的供应链和较低的加工成本。国外不少硬件项目在众筹网站超额完成融资，但美国制造业的“软肋”让这些项目迟迟无法从概念变成实物，这就给中国传统制造业重镇提供了无限遐想的可能。

7.7.3 创客中国的市场细分定位

对于小团队来说，一旦涉足大众消费级的产品，无异于九死一生。而去做一些大公司不会触碰的细分市场，只要产品能打入哪怕是非常细分的人群，数千台的销量也能让一家创业公司持续获得健康的现金流。以深圳华强北的一个只有 5 名成员的创业小团队为例，其研发的开发板 Cubieboard 年产只有 2 万台，年销售却可超过 100 万美元，利润也有 60 万美元，5 个团队成员平分，每人每年能有过百万的收入，而且都是刚毕业不久的年轻人。

创客们确实可以靠高利润的产品获得不错的收益，而且无须着眼于大众级别的产品。只要抓好愿意买单的少数细分人群就足够了，为什么还要走薄利多销以量取胜的老路呢？在即将到来的软硬结合的大时代，创客们可以在各自的小地盘上成立一千个小而美的公司，而不是杀入到只剩一个大公司的红海市场中成为炮灰。对于创客来说，传统 VC（Venture Capital, 风险投资）的模式未必行得通，因为 VC 们一般偏爱于大众级别的产品。创客完全可以通过 Kickstarter 这类众筹平台募得启动资金，然后找 Sseed Studio 这样的平台做小批量的生产，年产数千件，利润超过 30%，则足以让团队活下去，在现金流稳定增长几年之后再伺机切入更大的市场。

同时，随着开源文化的兴起，山寨手机的神话由此破灭，使得那些眼界变远的“厂二代”（山寨厂商的子辈接班人）愿意对新的技术趋势做出积极应变。如果说他们打江山的父辈靠的是规模化和薄利多销，“厂二代”们要想寻求山寨厂商的转型则需要寻找差异化、小批量和细分人群，这恰恰是可以和创客们对接的。未来一些开明的山寨厂商甚至会主动与创客群体对话，在创客中征求创意和设计外包，靠着细分和差异化完成转型。

创客们要想在大公司的眼皮底下寻找到面向细分市场的产品，很大程度上要依赖于多学科知识的交叉融合。正因为要涉及多学科的知识，使得大公司的研发成本很高，而产品的市场规模又不大，所以只能选择放弃。“要做 3D 打印机，需要机械工程师、写程序的人、懂电子的人、做工业设计的人、不同跨学科跨领域的人合力，并且需要开放的空间，创新的东西才会源源不断。”上海新车间创始人李大维谈到。同样，在北京创客空间创始人王盛林看来，创客的本质就应该是一种跨界的思维碰撞，不同背景的人取长补短，才能产生非常规的创新，创造出大公司无法低成本研发的细分产品来。

目前全球已有一千多家创客空间。自 2010 年从第一个国内创客空间在上海开办以来，国内创客文化得到了迅速的发展。北京、上海、深圳、杭州等几个大城市先后诞生了创客空间，以北京的创客空间、上海的新车间、深圳的柴火空间等为代表。其中，上海政府更提出要在“十二五”期间建立 100 个科技创新屋的目标。此外，还有面向全国范围的综合性创客网站平台“创客中国”（<http://www.makerchina.net>），对开源软件 Android、开源硬件 Arduino、视觉艺术，到具体的创意实现（如制作四轴飞行器、魔方机器人等）都有专门的学习讨论板块。这些网络平台上汇集的创客爱好者，在多学科的碰撞中产生灵感和火花，呈现星星之火可以燎原之势，将来中国的乔布斯和比尔·盖茨必将从他们之中诞生。

第8章

创客实战：四轴飞行器

《冬夜读书示子聿》（宋代·陆游）有云：“纸上得来终觉浅，绝知此事要躬行”。陆放翁的这句话可谓点中了创客的本质。如今是信息大爆炸的时代，除了汗牛充栋的各类书籍之外，网络上的资料和论文更是数不胜数。然而，如果你不实际动手做，对事物的理解就永远停留在纸面上，里面的技术细节和技巧无从知晓。因此，发挥你的执行力，立即找一个感兴趣的点开始做起来！在做的过程中你就会发现：原本看起来那么多、那么繁杂的知识，竟可以被有条理地串成一条线。接着，你变得能够看懂深奥的理论并把它们付诸试验，于是一条条纵横交织的相关线索开始完善你的知识面，让你的思维纵横驰骋。随着研究的进一步深入，你甚至形成了自己的知识体系，乃至可以写书立著。因此，“实践是检验真理的唯一标准”。

在本章中，我们教你如何结合 3D 打印和智能算法，去亲自动手做一架四轴飞行器。虽然四轴飞行器的出现时间并不长，但已成为创客们喜爱的热门制作项目之一。翱翔蓝天是每一个人生都会有的理想，虽然我们一时半会成为不了钢铁侠，但看着自己亲手制作的无人机飞行在天空，也是人生的一大乐趣。

8.1 你准备好了吗：自己制作四轴飞行器

相对于载人直升机体积大、成本高，在城市、山区等复杂环境中飞行时存在安全风险，无人机体积小、成本低，飞行机动性强，具有显著优势。四轴飞行器就是一种具有 4 个对称旋翼的无人直升机，具有垂直起降、结构简单、操作方便及机动灵活等优点，正越来越受到航模爱好者的热衷追捧。目前市面上也已经有了多款产品在售，代表性的产品有 Parrot AR.Drone 2.0、DJI Phantom 等。

下面以 Parrot AR.Drone 2.0 为例做一介绍，如图 8-1 所示，这款飞行器基于四轴智能设计，拥有 4 个独立螺旋桨。这款飞行器的下方还加装有重力感应装置、陀螺仪、机械控制芯片等部件，利用智能飞行技术可以纠正风力和其他环境误差。飞行器在室内外均可使用。如图 8-1 左边所示，室外可去除防护罩，以防止侧风影响稳定性；若飞进室内，可以把全壳护罩带上，防止碰撞损坏，如图 8-1 右边所示。



图 8-1 Parrot AR.Drone 2.0 外形。左：无防护罩；右：有防护罩（图片来源：Parrot）

Parrot AR.Drone 2.0 内置 ARM9 CPU 和 Linux 操作系统，通过无线 Wi-Fi 来控制，因此用户可以使用 iPhone 手机、iPad、iPod Touch、Android 手机或平板电脑等对其进行飞行控制操作。随机搭载了两台摄像机，一台在前面（1 280×720 像素），一台在下面（对着地面）。视频画面会同步传送到手机或平板电脑上。此外，两台以上 Parrot AR.Drone 2.0 可模拟空战游戏，而玩家只需要到苹果 App Store 中下载相关软件即可，如图 8-2 所示。



图 8-2 通过 iPhone 手机的 Wi-Fi 来遥控四轴飞行器

Parrot AR.Drone 2.0 虽好，但它却不是开源的，也就是说电路图纸、机械设计图纸，还有源代码等都不公开。这显然不能满足创客的基本需求，因为我们经常要进行二次开发，要对其中的一些功能进行改进。所以，我们还是自己 DIY 一架吧！值得庆幸的是，现在已经有一些国内外的创客把自己的设计开源共享了，你无须从零开始。在本书的网络下载栏目中，有本章项目的完整图纸和 3D 模型，你直接用它们制作或者 3D 打印出来即可！

8.2 器件与3D打印

目前，越来越多的创客对亲手制作一架四轴飞行器情有独钟。一时间，有关 DIY 制作四轴

飞行器的帖子在网络上层出不穷，各种各样的资料也是接踵而至。由于四轴飞行器涉及很多方面的知识，初学者往往会对那些专业术语茫然失措、不知所云。为了让大家都清楚地了解四轴飞行器的飞行原理、结构构成、器件选用，本节将从以下几个方面对四轴飞行器做一个简单的介绍。

8.2.1 四轴飞行器 DIY 所需的器件汇总

根据四轴飞行器的复杂程度，在这里我们把所需器件分为：四轴飞行器的基本配置器件和四轴飞行器的高级配置器件。

四轴飞行器基本配置器件是组装一套飞行器的基本必需器件，也是飞行器能否正常飞行的关键。在这里我先简单地介绍一下四轴飞行器基本配置器件及数目要求，更详细的介绍将会在后文中逐渐展开。

- 无刷电机（4 个）
- 电子调速器（简称电调，4 个）
- 螺旋桨（4 个，2 个正桨，2 个反桨）
- 飞行控制板（1 套）
- 航模动力电池（若干组）
- 遥控器（1 个）
- 匹配遥控器的接收机（1 个）
- 机架（市场上有现成的机架，在本章中我们将用 3D 打印机自己设计加工）
- 平衡充电器（1 套）

四轴飞行器高级配置器件是为了让飞行器能够完成更高要求的飞行任务所需的器件，包括飞行器的定位、飞行器航拍、飞行器特技等。因为不同应用的飞行器所需的高级配置也各不相同，现在简要介绍几种常用的四轴飞行器高级配置器件。

- 航模专用高清微型摄像头。因为飞行器飞行高度的原因，要想比较清楚地看清四周的景物就必须选择像素比较高的摄像头。目前这种微型摄像头的市场价格从几十元到几百元，甚至几千元的都有。当然要是能买到能够调焦的摄像头更好，但是价格可不菲哦。
- FPV（First Person View，即“第一人称视角”）远距离无线图传模块。有了摄像头，要想完成对图像的实时传输还需要一套给力的图传模块。目前来说，现有的图传模块大多数工作在 5.8GHz 这个频率上。常见的图传模块的功率从 0.5W 到 2W 不等，而传输的距离从几十米到几百米的都有。
- 高精度 GPS 模块。有许多航模爱好者往往想通过对飞行器发送指令使飞行器能够固定在空间的某一点处悬停，达到“指哪飞哪”的效果。这就需要高精度 GPS 模块来帮忙了，它通过人为设定三维坐标轴，使得空间中的每一个点在这个坐标轴上都有特定的坐标，从而达到使飞行器定点飞行的目的。
- FPV 航模航拍显示屏。航拍显示屏的作用就是将接收到的图像数据通过显示屏显示出来，达到对航拍视频实时观测以及调整飞行器飞行路线的作用。选择显示屏的时候要注意和摄像头及传输模块搭配。另外，为了较为方便地观察飞行器航拍的效果，通常情况下许多航模爱好者往往把航拍显示屏和遥控器安装在一起。

8.2.2 四轴飞行器的遥控器和接收机

想要体验远距离控制飞行器的快感，一对好的遥控器和接收机是关键。目前市场上的遥控器有许多的技术参数，初学者可能往往不能完全明了，现在此做一介绍。

遥控器的通道：通道就是遥控器可以控制的动作路数，比如遥控器只能控制四轴上下飞，那么就是 1 个通道。四轴在控制过程中需要控制的动作路数有：上下、左右、前后、旋转，所以最低得是四通道遥控器。而七通道的遥控器控制路数还包括：3 个辅助通道，以便操作者能够自由地发挥。图 8-3 为“天地飞”七通道遥控器的正面各个部件的示意图，图 8-4 为“天地飞”七通道遥控器接收机各个通道示意图。



图 8-3 “天地飞”七通道遥控器的正面各个部件的示意图（图片来源：深圳天地飞）

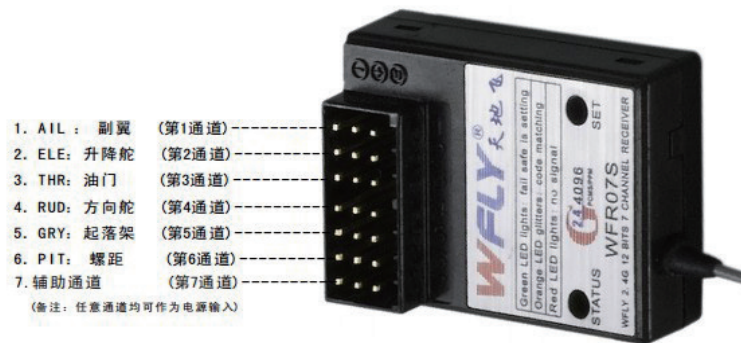


图 8-4 “天地飞”七通道遥控器接收机各个通道示意图（图片来源：深圳天地飞）

区分美国手和日本手：遥控器上油门的位置在右边是日本手，在左边是美国手。所谓遥控器油门，在四轴飞行器当中控制着供电电流大小，电流大，电动机转得快、飞得高、力量大，反之同理。判断遥控器的油门很简单，遥控器两个摇杆当中，上下扳动后不自动回到中间的那个就是油门摇杆。

遥控器的工作频率：通常市面上见到的遥控器频率大多数是 2.4GHz 的。因为在这个频率段中，和其他已用无线频道间隔较远，同时可用的频道比较多，不会因为多个遥控器同时控制而引起干扰。

8.2.3 四轴飞行器的飞行控制板

飞行控制板（简称“飞控”），顾名思义就是控制飞行器飞行过程的主板。如果没有飞控板，四轴飞行器就会因为安装、外界干扰、零件之间的不一致性等原因造成飞行力量不平衡，后果就是左右、上下地胡乱翻滚，根本无法飞行，飞控板的作用就是通过板上的陀螺仪、加速度传感器等器件对四轴飞行状态进行快速调整（都是瞬间的事，不要妄想用人肉完成）。当发现右边力量大，向左倾斜，那么就减弱右边电流输出，电机变慢，升力变小，自然就不再向左倾斜，反之亦然。常见有 KK、FF、玉兔等品牌（图 8-5 是常用的几种飞控板示例）。

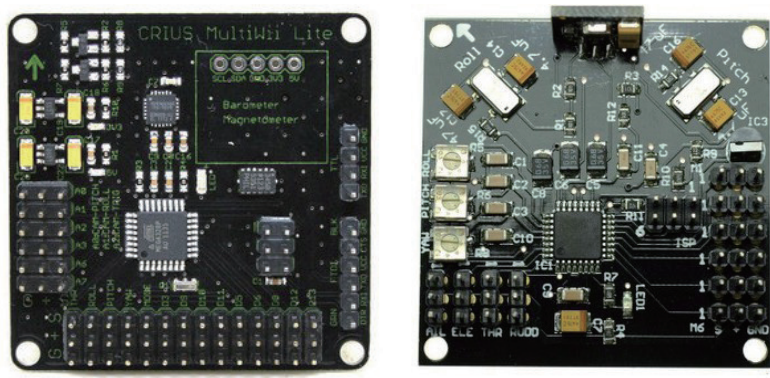


图 8-5 常用的几种飞控板。左：KK 飞控；右：MWC MultiWii Lite 轻量版 4 轴飞控

市场上的飞控类型虽然各式各样，但是基本上大多数飞控都分为几种飞行模式：经典 4 轴 + 模式、4 轴 × 模式、6 轴 + 模式、6 轴 × 模式、Y 型 4 轴、Y 型 6 轴、2 轴阿凡达飞行器等。而“经典 4 轴 + 模式”和“4 轴 × 模式”是四轴飞行器最基本的两种模式。现在就来具体说说这两种模式的区别。

对于四轴飞行器来说，所谓飞行模式，就是指飞行器飞行方向是否沿着飞行器的轴线，在这里通常把沿着飞行器相邻两轴线的对角线飞行的模式称为 × 模式（XCopter），而把沿着飞行器轴线方向的模式称为 + 模式（QuadCopter）。在买的时候需要向卖家说明你要的是哪种模式。一般来说，× 模式要难飞一点，但动作更灵活；+ 模式要好飞一点，但动作灵活性差一点，所以适合初学者。特别注意：× 模式和 + 模式的飞控安装是不同的，体现在电机旋转的方向上（如图 8-6 所示）。如果飞控板安装错误，飞行器会剧烈地晃动，根本无法起飞。

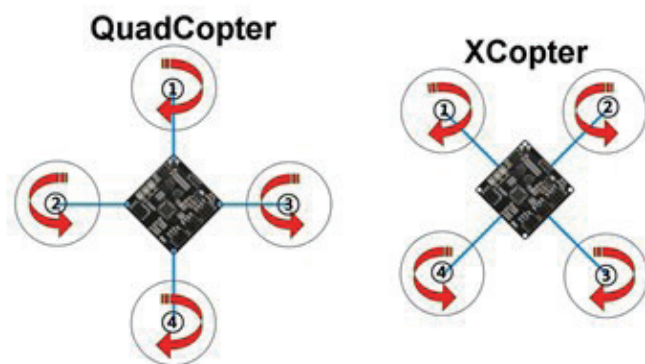


图 8-6 四轴飞行器中 + 模式和 × 模式的比较（图片来源：kkmulticopter）

8.2.4 四轴飞行器电调的选用

电调，全称电子调速器，作用就是将飞控板的控制信号转变为电流的大小，以控制电机的转速。因为驱动电机需要的电流是很大的，通常每个电机正常工作时，平均有 3A 左右的电流，如果没有电调的存在，飞控板根本无法承受这样大的电流（另外也没有驱动无刷电机的功能）。同时电调在四轴当中还充当了电压变化器的作用，将 11.1V 的电压变为 5V 为飞控板和遥控器供电。四轴常用的电调有：好盈、中特威、新西达等品牌（如图 8-7 所示是常用的几种电调示例）。



图 8-7 常用的几种电调。左：好盈 40A 电调；右：中特威 20A 电调

通常情况下，电调都会标上电调能够提供的最大电流数，单位为 A（安培），如 30A、40A 等。大电流的电调可以兼容用在小电流的地方，而小电流电调不能超标使用。为了保险起见，一般大家都会配置 30A 或 40A 电调（玩家用 20A 电调的也不少），买大一点，以后还可以用到其他地方去。但是电流越大的电调往往也越重，为了进一步减少飞行器的重量，建议选择电调时最好是测出自己的电机满油门最大电流，在此之上再增加 10% 左右的余量即可。

有时还会见到“四轴专用电调”这个名词。诚然，电调有快速响应和慢速响应的区别，四轴飞行要求电调快速响应。但大多数常见电调都是可以编程的，能通过编程来设置响应速度，所以其实并没有什么专用一说。

8.2.5 四轴飞行器的无刷电机和螺旋桨

生活中常用的电机分为有刷电机和无刷电机，四轴飞行器选用的是无刷电机。因为它力气大，耐用。常用的电机有：新西达、朗宇、银燕等品牌（图 8-8 是常用的几种无刷航模电机示例），不同品牌的电机都有不同的型号。经常有人说 2 212 电机、2 018 电机等，到底是什么意思呢？这其实是电机的尺寸。不管什么牌子的电机，具体都要对应 4 位这类数字，其中前面 2 位是电机转子的直径（单位 mm），后面 2 位是电机转子的高度。注意，不是外壳哦。简单来说，前面 2 位越大，电机越大，后面 2 位越大，电机越高。又高又大的电机，功率就更大，适合做大四轴。通常 2 212 电机是最常见的配置。



图 8-8 常用的无刷航模电机。左：银燕 BL2210，KV1560 电机；右：朗宇 500KV，X3520-06 电机

电机的另外一个重要参数就是电机的 KV 值。每个无刷电机都会标注多少 KV 值，这个 KV 是外加 1V 电压对应的每分钟空转转速，例如，1 000KV 电机，外加 1V 电压，电机空转时每分钟转 1 000 转，外加 2V 电压，电机每分钟空转就是 2 000 转了。注意，这些高速的无刷电机使用的是三相交流电，不像直流电机那样接上一定电压就可以旋转，还需要利用 8.2.4 节介绍的电调来接受单片机的 PWM 指令（详见本章 8.4.1 节），以获得无刷电机需要的高频交流电。

螺旋桨是指靠桨叶在空气或水中旋转，将发动机转动功率转化为推进力的装置。四轴飞行器为了抵消螺旋桨的自旋，相隔的桨旋转方向是不一样的，所以需要正反桨。正反桨的风都向下吹。适合顺时针旋转的是正桨，适合逆时针旋转的是反桨。安装的时候，一定记得无论正反桨，有字的一面是向上的（桨叶圆润的一面要和电机旋转方向一致）。同电机类似，桨也有 1 045、7 040 这些 4 位数字，前面 2 位代表桨的直径（单位：inch，1 inch = 25.4 mm），后面 2 位是桨的角度。如图 8-9 所示是常用的几种四轴飞行器正反桨。



图 8-9 常用的几种四轴飞行器正反桨

最后，谈一下电机与螺旋桨的搭配问题。螺旋桨越大，升力就越大，但对应需要更大的力量来驱动；螺旋桨转速越高，升力越大；而电机的 KV 越小，转动力量就越大。所以，大螺旋桨需要用低 KV 电机，小螺旋桨需要用高 KV 电机（因为需要用转速来弥补升力不足）。如果高 KV 带大桨，力量不够，转动就很困难，实际还是低速运转，电机和电调都很容易被烧掉。如果低 KV 带小桨，完全没有问题，但升力不够，可能造成无法起飞。

8.2.6 四轴飞行器的电池和充电器

根据锂离子电池所用电解质材料的不同，锂离子电池分为液态锂离子电池（Liquified Lithium-Ion Battery，简称为 LIB）和聚合物锂离子电池（Polymer Lithium-Ion Battery，简称为 PLB）。四轴飞行器常用的电池为聚合物锂电池，采用固体聚合物作为电解质。

大家在买电池的时候经常会看到电池上标有 $\times\times\times\times\text{mAh}$ ，这个表示的是电池容量，如 1 000mAh 电池，表示在 1 000mA（1A）的电流下，可以工作 1 小时；如果以 500mA 放电，则可以持续工作 2 小时。

电池后面的 $\times\times\text{C}$ ，例如 25C、30C 等，代表电池的放电能力，这是普通锂电池和动力锂电池最重要的区别，动力锂电池需要很大的电流放电，这个放电能力就用 C 来表示。如 1 000mAh 电池，标准为 5C，那么用 $5\times 1\,000\text{mAh}$ ，得出电池可以以 5 000mA 的电流强度放电。这很重要，航模一般都需要比较大的瞬间功率，如果用低 C 的电池，大电流放电，电池会迅速损坏，甚至自燃。

另外，在用锂电池的时候我们常常会发现：电池后面有 2S、3S、4S，这个则代表锂电池的节数，一节锂电池的标准电压为 3.7V，那么 2S 电池，就代表有 2 个 3.7V 电池在里面，电压为 7.4V，同理，3S 是 11.1V，4S 是 14.8V 等。在大多数情况下，四轴飞行器所用的电池常见的有：7.4V（2S）、11.1V（3S）、14.8V（4S），电池的品牌有：花牌、自主 OM、格式 /ACE 等，如图 8-10 所示。



图 8-10 左：1550mAh 11.1V 25C 航模电池组；右：11.1V 2200mAh 25C 航模电池组

在这里说说四轴飞行器电池的大致选用原则。总地来说，这与选择的电机、螺旋桨，想要的飞行时间都是相关的。因为电池的容量越大，C 越高，S 越多，但是随之而来的是电池的重量提高。如果飞行器体积较大，且是用大桨，因为整体搭配下来功率高，自身升力大，为了保证可玩时间，可选高容量，高 C 值，3S 以上电池。最低建议 1 500mAh、20C、3S。而小四轴，因为自身升力有限，整体功率也不高，就可以考虑小容量，小 C 值、3S 以下电池。

一般来说，为了保护四轴飞行器所用的聚合物锂电池，延长电池的使用寿命，我们选用平衡充电器，比如 B6 充电器。何为平衡充电器？在这里为了说明问题，我们举个例子，例如，对于 3S 电池，内部是 3 个锂电池。因为制造工艺原因，没办法保证每个电池完全一致，充电、放电特性都有差异，电池在串联的情况下，容易造成某些放电过度或充电过度，而其他一些又不饱满等，解决办法是分别对内部单节电池充电。因此，在上面的聚合物锂电池的图中我们可以看到：动力锂电池都有两组线，一组是比较粗的输出线（2 根），另一组是单节锂电池引出的比较细的线（与电池的 S 数有关，3S 就是 3 根）。

8.2.7 四轴飞行器的连接线选用

由于四轴飞行器所用的是无刷电机，在正常工作时，工作电流通常是几安到十几安，甚至是几十安。在这样大的电流环境下，普通导线所能允许的电流强度往往不能满足飞行器的需要，所以我们必须根据自己的需求选择合适的线材，这时就需要对线材的一些参数有所了解。目前各种网站上的航模店所卖的线材几乎都标注了 AWG 是多少，这个 AWG 其实是一种导线的标准，AWG (American Wire Gauge) 即“美国线规”。通过查寻导线参数就可以知道这个导线适用的环境。AWG 一般用“AWG+ 数字”来进行标示，后面的数字就是参数，例如 AWG12。

我们可以发现一个简单的规律就是，参数越小，导线越粗，比如，AWG12 肯定比 AWG18 粗，所以在设计飞行器选择线材时千万要注意识别线材的规格，不要买错了。

下面我们来谈谈导线的最大电流原理。在航模领域，我们买不同粗细导线的理由无外乎是为了能过更大的电流（其他领域可能还考虑强度、外部环境等因素）。何为最大电流？即中间铜芯和绝缘皮融化之前的电流。请注意，这里的绝缘皮也很重要，航模常用高温线，那是因为环境温度高、电流大，这里的高温其实是指绝缘皮，铜芯本身熔点是固定的。如果不用高温线，即使铜芯没问题，绝缘皮融化了也会短路，造成电路损坏。鉴别高温线很简单，用 25W 左右烙铁烫一下，如果一下了就融化了则肯定不是。

8.2.8 四轴飞行器机架的 3D 打印

现在不同的航模爱好者做出的四轴飞行器各式各样，并没有一定的规范。从理论上讲，只要 4 个螺旋桨不打架就可以了，但要考虑螺旋桨之间因为旋转产生的乱流互相影响，建议还是不要太近，否则影响效率。这也是四轴用 2 叶螺旋桨比用 3 叶螺旋桨多的原因之一。另外，3 叶桨还有个缺点，平衡性比较差，不易控制。所以要想真正做出一个比较稳定并且适用的四轴飞行器机架，还需要对设计不断调整和改进，使之性能达到最优。

国外一名叫 Adam Polak 的创客，在他的网站上（<http://polakiumengineering.org>）发布了其完成的新项目：超级酷的微四轴飞行器 R2。机架的 3D 打印图纸请到这个链接下载：<http://www.thingiverse.com/thing:29632>。这架小型四轴飞行器的零件采用 ABS 材料进行 3D 打印。轴很坚固，施加重力时能够弯曲而不至于断裂。

如图 8-11 所示，完整的机架包括 4 个轴和 2 块架板，可用个人 3D 打印机打印出来。



图 8-11 完整的机架包括 4 个轴和 2 块架板（利用个人 3D 打印机制作）（图片来源：Adam Polak）

组装时，建议使用中等稠度的 CA 胶水与 CA 促进剂（Accelerator，一种提高黏合反应速度的用量较少的物质）进行黏合。首先把 CA 胶水涂在顶板的底面，促进剂涂在轴的正面黏合处。将轴固定到相应位置，直至黏合。

然后把 CA 胶水涂于底板的正面，促进剂涂在每个轴脚上。接着将已经安装好的部分固定到底板上，确保轴脚都固定。

机架完成后，就可以开始安装电路板了。四轴飞行器的组装效果如图 8-12 所示。

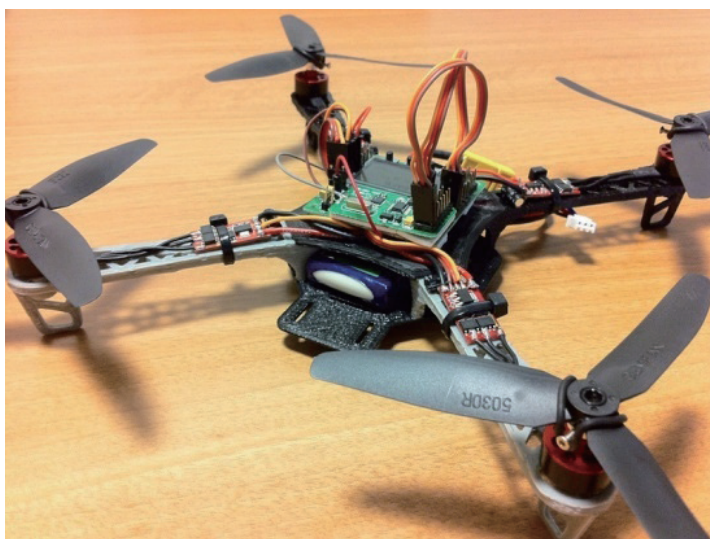


图 8-12 组装好机架和电路板的四轴飞行器

8.3 三轴陀螺仪和加速度计的入门与调试

绕一个支点高速转动的刚体称为陀螺（Top）。在一定的初始条件和一定的外力矩的作用下，陀螺会在不停自转的同时，还绕着另一个固定的转轴不停地旋转，这就是陀螺的旋进

(Precession), 又称为回转效应 (Gyroscopic Effect)。陀螺进是日常生活中常见的现象, 许多人小时候都玩过的陀螺就是一例。

根据物理学知识我们知道, 一个旋转物体的旋转轴所指的方向在不受外力影响时, 是不会改变的。人们根据这个道理, 用它来保持方向, 制造出来的东西就叫陀螺仪 (Gyroscope)。我们骑自行车其实也是利用了这个原理。轮子转得越快越不容易倒, 因为车轴有一股保持水平的力量。陀螺仪在工作时要给它一个力, 使它快速旋转起来, 一般能达到每分钟几十万转, 可以工作很长时间。

现代陀螺仪是一种能够精确地确定运动物体的方位的仪器, 它是现代航空、航海、航天和国防工业中广泛使用的一种惯性导航仪器。根据需要, 陀螺仪器能提供准确的方位、水平、位置、速度和加速度等信号, 以便驾驶员用自动导航仪来控制飞机、舰船或航天飞机按一定的航线飞行, 而在导弹、卫星运载器或空间探测火箭等航行体的制导中, 则直接利用这些信号完成航行体的姿态控制和轨道控制。作为稳定器, 陀螺仪器能使列车在单轨上行驶, 能减小船舶在风浪中的摇摆, 能使安装在飞机或卫星上的照相机相对地面稳定等。

对于四轴飞行器而言, 它必须要通过陀螺仪来“感知”自己的姿态, 然后通过自己的“大脑”——控制系统来调节自身平衡, 以达到稳定飞行的目的。但是在陀螺仪的实际调试中, 我们会发现, MEMS (Microelectromechanical Systems, 微型机电系统) 陀螺仪自身的性质决定了 MEMS 陀螺仪是存在温度漂移和零点漂移的, 所以在实际的操作中, 仅仅依靠 MEMS 陀螺仪来感知自身姿态是不可靠的。

图 8-13 描述了由于漂移的影响由陀螺仪积分得到的角度和实际角度的曲线 (红线代表陀螺仪积分得到的角度, 黑线代表实际角度), 可以看出红线明显偏离了黑线。

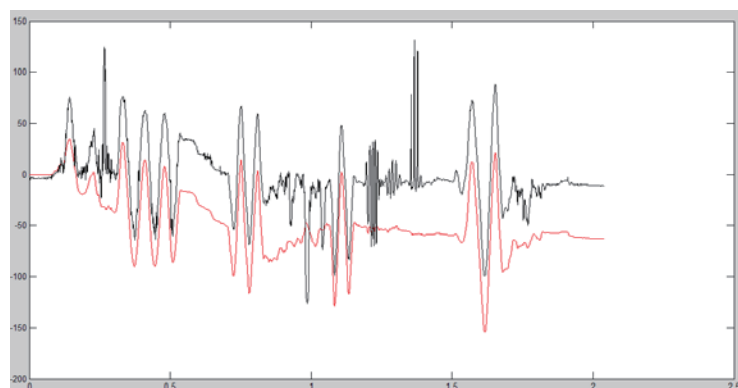


图 8-13 由陀螺仪积分得到的角度和实际角度的曲线

因此, 为了解决这一问题, 我们引入了加速度计 (又叫重力感应器)。加速度计可以测量加速度, 包括重力加速度。在静止或匀速运动的时候, 加速度计仅仅测量的是重力加速度, 而重力加速度与地球坐标系是固联的, 通过这种关系, 可以得到加速度计所在平面与地面的角度关系。

和陀螺仪相比, 加速度计测量的是绝对的“倾角”, 而陀螺仪测量的是“倾角的变化速度”。有人可能觉得既然陀螺仪有漂移, 那么只用加速度计就好了, 因为它可以测量绝对角度。但是请

注意，加速度计获得的绝对倾角必须是在物体静止或者匀速的情况下测量的，飞行器在运动过程中是测不准的。所以一般需要两个传感器配合：陀螺仪用来测变化量，加速度计用来修正误差。

接下来的工作就是将加速度计传感器和陀螺仪传感器的测量数据进行数据融合。数据融合的方法有很多种，比如四元数法、卡尔曼滤波方法、耦合滤波方法等。各个算法具体的原理和代码这里就不做详细介绍了。为了使融合的概念更为直观，在这里我们用图 8-14 来说明。

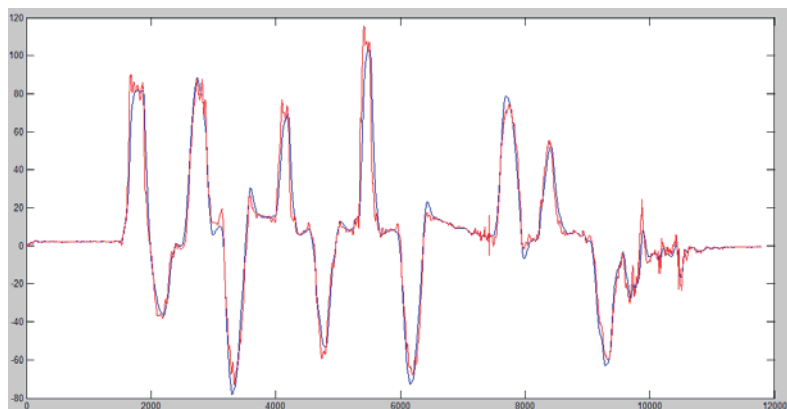


图 8-14 应用融合算法得到的某个轴角度

从图 8-14 中（红色为实际角度，蓝色为融合角度），我们可以明显看出耦合角度和实际角度的跟随性比较好，同时又有一定的低通特性，可以滤除一定频率的抖动，这对于飞行器的控制是至关重要的。

通过上面的分析和融合算法我们就可以得到四轴飞行器的真实姿态角，有了姿态角对于四轴飞行器的控制来说就得心应手了。

下面为大家介绍一些陀螺仪和加速度计的生产商。

陀螺仪生产商

- 日本 Silicon Sensing：代表产品 CRS03，性能非常优越，价格略高，主要应用于无人机和惯导系统中。
- 奥地利 SensorDynamics：多数为 SSOP 封装，焊接比较方便，通信接口为 SPI 接口。
- 美国 Systron Donner Inertial：代表产品 LCG50，主要应用于无人机和惯导系统。
- 美国 ADI：代表产品众多就不一一列举了，主要应用中端及低端市场，近些年也出高端传感器，但价格略高。
- 芬兰 VTI：最近两年才出了 SCC1 300（单轴陀螺仪三轴加速度传感器），价格比较贵。

加速度传感器生产商

- 美国 Silicon Designs：代表产品 Model 系列，主要应用于无人机和惯导系统中。
- 美国 ADI：代表产品众多就不一一列举了，主要应用中端及低端市场，近些年也出高端传感器，但价格略高。
- 芬兰 VTI：代表产品 SCA100T、SCA103T 等倾角传感器，零偏数据和长期稳定性及重复性好，主要应用于倾角测量（只适合静态或准静态环境）。

- 瑞士 Colibrys：代表产品 MS7 000、MS8 000、MS9 000，是高带宽、高精度的电容式加速度计，主要应用于无人机和惯导系统。
- 奥地利 SensorDynamics：多数为 SSOP 封装，焊接比较方便，通信接口为 SPI 接口。

8.4 自制基于Arduino的飞控板

众所周知，任何一款飞行器都要有一整套完备的电控系统来控制，这样飞行器才能够正常、稳定飞行。当然，四轴飞行器也不例外。下面，我们就从四轴飞行器电控系统的基本组成单元开始，逐步了解各模块的功能以及整个电控系统结构。

8.4.1 四轴飞行器的基本电控结构

为了方便起见，我们用表格的形式（见表 8-1）列出了所需要的主要元器件名称、数量，并且简要介绍了它们各自的功能。

表 8-1 四轴飞行器所需的电控元器件及功能

名称	数量 / 个	功能
无刷电机	4	带动桨叶转动
电子调速器	4	控制电机的转速
飞行控制板	1	稳定四轴飞行器飞行
遥控器	1	操控四轴飞行器飞行状态
电池	1	提供电源
充电器	1	给电池充电

其中，遥控器、电池、充电器的功能顾名思义，不再赘述。你可能对剩下的 3 个单元模块不太了解，下面我们做进一步的说明。

无刷电机：相比于普通直流电机，去除了电刷，因此也就没有了电刷与轴承摩擦产生的火花，减少了电火花对遥控无线设备的影响。同时，摩擦的减小，也会让电机运行更顺畅、噪声更低、寿命更长。总之，无刷电机转速快、控制精度高，满足四轴飞行器要求。此外，无刷直流电机一般使用的都是三相交流电，有 3 根接线，如图 8-15 左边所示。

电子调速器：简称电调，它有 8 根接线：2 根输入电源线，和电池连接；3 根控制线，和单片机连接；3 根输出电源线，和电机连接。在所有的 3 根控制线中，一根是地线，一根是引出的 5V 电压线（这是个贴心的设计，Arduino 将从这里取电，不用再考虑电池的问题），最后一根则是控制线，接收 PWM 信号。此外，电调有个重要的参数叫最大电流，一般选择 20A 以上的，如图 8-15 右边所示是一个最大电流 150A 的电调。

飞行控制板：简称飞控板，是电控系统的核心部分，它的主要作用就是计算陀螺仪采集到的空间信息，然后对四轴飞行状态进行快速调整。关于飞控板，我们已经在本章 8.2.3 节中进行了介绍，在此不再赘述。

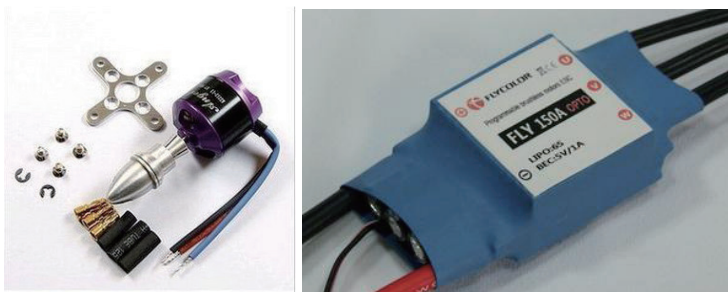


图 8-15 左：航模无刷电机；右：电子调速器（图片来源：深圳市飞盈佳乐）

从上面的介绍中，我们可以看出，除去 2 根输入电源线之外，电调是一个有 2 个输入端和 3 个输出端的元件，而无刷电机正好也是 3 根接线！原来电子调速器的 3 根输出端就是要和无刷电机的 3 个输入线相连接的，事实上，电子调速器就像一个桥梁，实现了飞控板与无刷电机的连接，从而使飞控板可以控制无刷电机的转动大小以及方向，进而实现对飞行器姿态的控制。图 8-16 画出了一般四轴飞行器电池、电调、电机、飞控板的布局以及电路连接关系。其中，飞控板与电调之间的红线表示连接电池电源线正极，黑线表示连接电池电源线负极，绿线以及紫线表示与控制板（我们将选用 Arduino）相连的电源线，橘色线表示 PWM 波控制线。

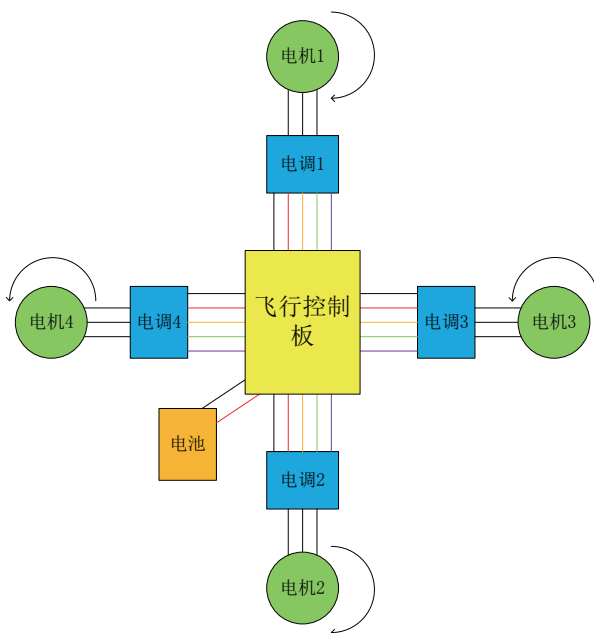


图 8-16 四轴飞行器电控系统连接示意图

可能有些读者对 PWM 波还不太了解，在此我们做一些简要的介绍。PWM 的全称是“Pulse-Width Modulation”，直接翻译过来就是“脉冲宽度调制”的意思，也就是说，PWM 波是控制脉冲宽度的。我们知道，高电平可以让 LED 灯发亮，而低电平可以让 LED 灯熄灭，但是如何让 LED 灯介于明暗之间呢？我们可以设想，是不是高低电平快速变化（超过人的反应速度）就会让人感觉亮度既不是最亮也不是最暗呢？猜对了，这就是 PWM 波的一个用处，它可以让元件产

生处于最大值与最小值之间的任意效果，只需调节 PWM 波最大值与最小值之间的时间比，即占空比，如图 8-17 所示即为一些不同占空比的 PWM 波。

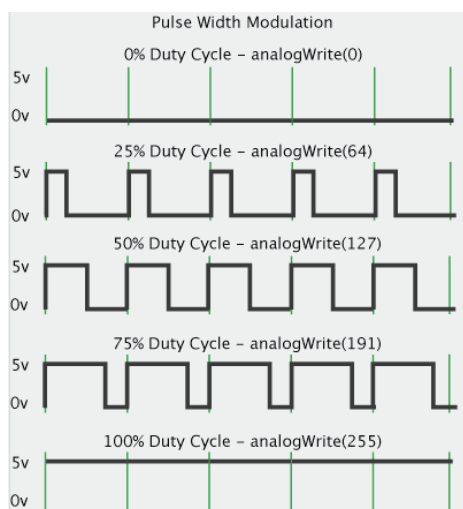


图 8-17 不同占空比的 PWM 波（图片来源：Arduino）

对于我们的四轴飞行器项目来说，不同占空比下的 PWM 波可以用来控制无刷电机的转速大小。对于 Arduino 来说，只有数字引脚区的 3、5、6、9、10、11 这 6 个引脚可用来输出 PWM 波，虽然数目不多，但是足够分配给 4 个电调来用。

通过本节介绍，我们了解了四轴飞行器的基本元件、模块功能以及元件布局，并且可以得出这样的结论：在四轴飞行器项目的电控系统中，除了选择电池、电机、电调、陀螺仪以及加速度计的选型之外，成败的关键在于飞控板的设计。因此，结合第 7 章所学的关于 Arduino 的知识，我们将在下一节学习基于 Arduino 平台的飞控板制作。

8.4.2 飞行控制板的制作

四轴飞行器相对于常规航模来说，最复杂的就是电子部分了。之所以能飞行得很稳定，全靠电子控制部件对四轴飞行器的飞行状态进行快速调整。从 8.4.1 节的学习中，我们知道飞控板就是电控部分的核心模块，但是我们没有仔细分析它的组成。一般来讲，飞控板由传感器、控制器、电机驱动模块和通信模块 4 个部分组成，接下来我们将一一介绍。

在飞控板上，最重要的两个传感器即为陀螺仪和加速度计。在常规固定翼飞机上，陀螺仪并非常用器件。然而，四轴飞行器则必须配备陀螺仪，否则就无法飞行，更谈不上稳定飞行。不但要有，还得是 3 个轴向（X、Y、Z 轴）都得有，这是四轴飞行器的机械结构以及动力组成特性所决定的。此外，在此基础上还要再辅以 3 轴加速度传感器，这 6 个自由度就共同组成了四轴飞行器飞行状态参数的基本部分。

当传感器采集到四轴飞行器的飞行状态参数之后，我们就需要对这些参数进行分析，并做出控制决策，进而校正四轴飞行器的状态以保持其稳定飞行。正如我们在第 7 章所看到的，Arduino 拥有强大的功能以及开源、易操作等诸多优势，完全适用于对陀螺仪、加速度计传感器

采集到的参数进行运算。还有一点特别重要的是，许多开源的四轴飞行器项目都是基于 Arduino 的，也就是说我们可以借鉴许多其他人的经验来开发自己的四轴项目，这将大大降低我们的工作量。综上所述，我们选用 Arduino 作为飞控板的控制器。

具体来说，为了简化驱动电机的代码工作量而充分利用开源资源，我们选择使用 Arduino Motor 扩展板，该类扩展板直接插在 Arduino 开发板上便可完成电路连接，使用起来十分方便。如图 8-18 所示的扩展板即为一种可同时驱动 4 路直流电机的 Arduino Motor 扩展板，有 4 路 PWM 调速，可以满足我们的四轴飞行器项目要求。当然，有得必有失，扩展板的增加虽然方便了我们项目的开发，但也势必会增加飞控板的总体积以及重量，如果这些方面有严格要求，则需要采用其他方案，这里就不再展开介绍。

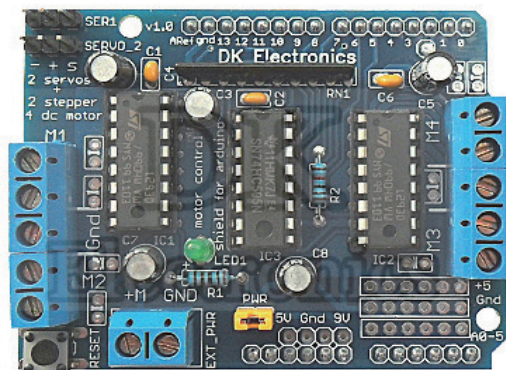


图 8-18 Arduino Motor 扩展板

在通信模块的设计中，可以选择蓝牙或 Wi-Fi，但是考虑到蓝牙信号传输距离太小（一般传输距离为 10m 左右），而四轴飞行器飞行实用性与飞行距离密切相关，所以我们选择传输距离较大的 Wi-Fi 信号来进行通信。幸运的是，许多智能手机带有 Wi-Fi 功能，并且若该手机的操作系统是 Android 系统，这将很大程度方便我们对四轴飞行器的控制——我们的 Android 手机就是遥控器。我们为什么要选择 Android 手机呢？因为 Android 系统是开源的并且操作界面友好，还有许多类似的开源项目可供我们借鉴，因此非常适合于我们的四轴飞行器项目。而在电路连接方面，我们仅仅需要做的就是给 Arduino 开发板插一个如图 8-19 所示的 Wi-Fi 扩展板。



图 8-19 Arduino Wi-Fi 扩展板

上面介绍了飞控板制作所需的传感器、控制器、电机驱动模块以及通信模块，现在我们将这几个模块插在一起，则飞控板的制作就大功告成了。

当飞控板制作完毕以后，即可按照图 8-16 所示的电路系统连接图将所有元件、模块连接起来，然后把它们固定在 8.2.8 节中 3D 打印好的机架上，并将桨叶安装在电机上面（注意电机 1、2 以及电机 3、4 的桨叶方向，避免安装错误），此时我们的四轴飞行器就设计完毕。然而，光靠这些机械器件以及电控元件四轴飞行器是不能飞行的，我们需要编写相关的飞行控制代码烧写在 Arduino 中。其实，在编写飞行控制代码之前，我们首先还需要开辟一条通道，以便四轴飞行器能够及时“听懂”我们的控制意愿。由于我们将选择使用 Android 手机作为遥控器来控制四轴飞行器，所以 Android 手机与四轴飞行器之间必须有一条畅通的通信路径。显然，Android 手机上的 Wi-Fi 模块与飞控板上的 Wi-Fi 模块就是为此服务的，具体的 Wi-Fi 通信过程我们将在 8.5 节学习。

8.5 遥控开始：Android手机的Wi-Fi通信

我们可直接利用手机、平板电脑等工具来遥控飞行器的飞行方向和姿态。到底它们究竟是怎样工作的呢？现在我就给大家介绍一种基于 Android 系统的 Wi-Fi 无线控制系统的基本工作原理，让大家更深入地了解其中的奥秘。

Android（详见第 7 章 7.5 节）是一款当前最为流行的手机操作系统，属于一个全开放的平台，因此开发者可以得到整个系统的源代码，并能对其进行修改。Android 系统是在 Linux 系统的基础上，经过了层层封装，最终提供给开发者的是大量的 Java API，在这里被叫作 Android API，于是，开发者就可以像开发一般的 Java 程序那样开发 Android 应用程序，这样的设计不仅降低了开发 Android 应用程序的难度，还增加了 Android 系统的界面友好度。

和一般的操作系统一样，Android 也是对硬件进行了多层的封装，使得应用程序的开发者和用户能轻松地操作硬件，完成他们所希望完成的事情。Android 中 Wi-Fi 驱动程序被编译成内核的模块，我们购买手机时，生产商已经帮我们做好了驱动。所以我们只需通过应用程序设置开关它即可，具体来说就是“Settings”→“Wireless & networks”→“Wi-Fi”。

接下来就是与 Wi-Fi 模块的通信了。我们可建立 TCP 服务器来进行连接，并创建通信协议来进行通信。由于 Wi-Fi 本身的通信非常可靠，因此我们只需自定义一个简单的通信协议，来控制四轴飞行器的上下、左右、前后、旋转、油门这几个操作即可。

这里介绍国内两位创客 lambo、whcf 所提供的方案。由于四轴飞行器的载重有限，因此我们选用轻质的 Wi-Fi 模块，市面上有的模块（如 ST-MW-09S）已经做到了，仅为 2g 重，如图 8-20 所示。

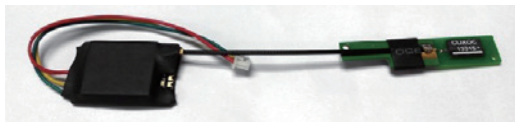


图 8-20 Wi-Fi 模块（ST-MW-09S）与天线模块（图片来源：密友电子）

在 TCP Server 模式下，用 Wi-Fi 模块监听设置的本机端口，如图 8-21 所示，有连接请求时创建连接，根据接收到的指令对四轴飞行器进行各种控制。

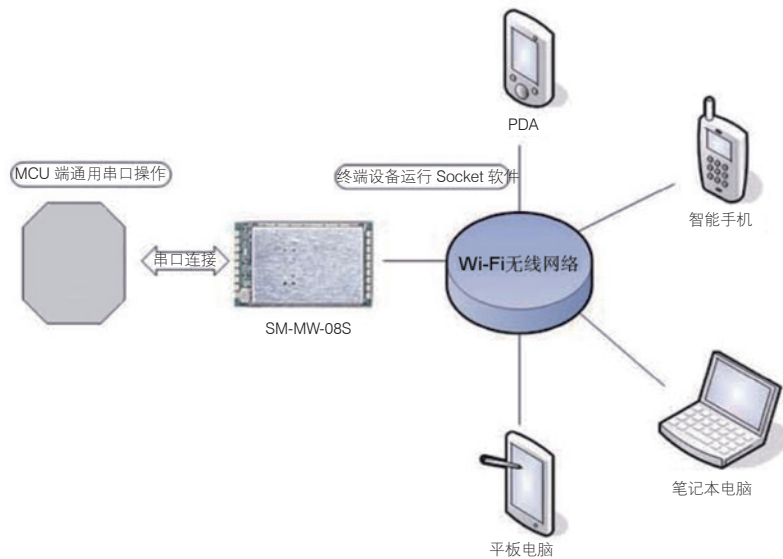


图 8-21 TCP Server 模式下，用 Wi-Fi 模块监听设置的本机端口

用户在用手机操作时其实非常简单，首先连接搭载在四轴飞行器上的 Wi-Fi 网络，如图 8-22 所示。



图 8-22 用手机连接搭载在四轴飞行器上的 Wi-Fi 网络

连接成功后，打开在手机上安装的飞行器控制 App，然后就可以通过手机触摸屏界面（如图 8-23 所示，类似于传统飞行遥控器）对四轴飞行器进行 Wi-Fi 无线遥控了。

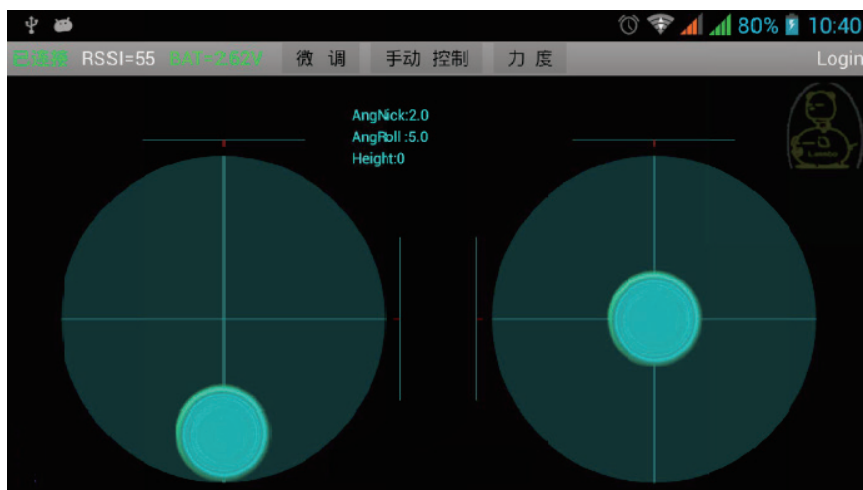


图 8-23 手机端控制程序（图片来源：lammbo）

本书的网络下载栏目中包含了 Wi-Fi 通信电子图纸和开源程序代码。此外，目前国内有一个四轴飞行器的开源项目 flexbot 可供参考，资源下载可访问 <http://www.flexbot.cc/>。

8.6 四轴飞行器的智能视觉跟踪

在 8.5 节中我们通过 Android（安卓）手机来手动遥控飞行器，确实很酷！然而在搞自动化学科的专业人士眼里，还不够酷哦！他们会问：你的飞行器升空后，可以自动跟踪地面上的目标吗？比如自动跟踪一辆红色轿车，或者某一位你心仪的女生？

航拍地面目标跟踪是指在低空飞行的无人机（比如我们制作的四轴飞行器）上安装摄像机，利用视觉跟踪获得地面目标的运动参数，以此控制云台旋转角度调整摄像机的姿态，使目标保持在视野中，并控制无人机跟踪目标飞行，如图 8-24 所示。地面目标跟踪是无人机视觉导航的重要内容之一，例如，在对地打击、城市反恐作战中，对车辆和人员的跟踪以及在海上搜救中对随波漂流人员的跟踪等。



图 8-24 在捕食者等军事无人机上都安装有云台，以调整摄像机的姿态（图片来源：presstv）

在对目标进行跟踪之前，我们首先需要对目标进行表示或者建模。而为了表示目标，则需要确定用什么特征来描述目标（可参考第6章6.2.1节“个性特征的描述与检测”）。常用的特征有颜色、边缘、轮廓、纹理、关键点等。然后，利用这些特征来构造目标的表示模型，常见的表示模型有：各种色彩空间的直方图、模板和模型（如骨架模型、AAM）、高斯混合模型、核密度估计、超像素等。

有了目标的表示模型，我们就可以开始目标跟踪了，代表性算法大致分为以下几种。

- 基于模板匹配的算法。比如 Mean Shift 算法是一种基于核密度估计的模式匹配算法，该算法的早期版本只能找到局部最大值，且传统的 Mean Shift 算法不能很好地处理遮挡问题以及不能自适应跟踪目标的形状、方向等。其后，有研究人员对其做了改进，比如 CamShift，就可以自适应物体的大小、方向，具有较好的跟踪效果。该算法最大的特点是跟踪速度快，在实际中应用较多。此外，还有基于子空间的方法，首先学习目标的特征子空间，然后根据当前帧的候选目标在特征空间的表示进行目标定位。基于稀疏表达模型（Sparse Representation）的跟踪方法可以看作是子空间跟踪的一个推广，它用一组目标模板作为字典。
- 基于概率预测的方法。依据贝叶斯估计理论，目标跟踪问题可以看作目标状态（包括目标的位置、大小、速度等）的估计问题，并通过递归贝叶斯估计获得状态的最大后验概率（MAP）。例如，Kalman 滤波假设物体的运动模型服从高斯概率模型，来对目标状态进行预测，然后与观察模型进行比较，根据两者之间的误差来寻找运动目标的状态。但是该算法的精度往往不高，因为高斯运动模型在现实生活中很多情况下得不到满足。粒子滤波克服了 Kalman 滤波的不足。它每次通过实验可以重采样粒子的分布，根据该分布对粒子进行扩散，通过扩散的结果来观察目标的状态，最后更新目标的状态。该算法因其能实现稳定跟踪而在实际应用中越来越多。
- 基于分类检测的跟踪算法：将跟踪问题视为区分特定目标（正样本）和背景（负样本）的分类检测问题。当训练样本足够多时，我们可利用第6章6.11.1节的 AdaBoost 方法对目标（如人脸）进行分类，此外还有 SVM、决策树、随机森林等分类器。基于分类检测的方法分为离线方法和在线方法。前者的检测器一旦初始化就不再变化，而后者利用当前帧的数据对检测器进行更新，从而对目标的变化有一定的自适应能力。

上面对目标跟踪算法的简介，相信普通读者看得云里雾里的吧？OK，没关系，我们这里就拿其中的两种算法进行详细解释，相信你一定能看得明明白白！

8.6.1 基于粒子滤波的目标跟踪算法

我们首先看一下到底什么是基于**粒子滤波（Particle Filtering）**的目标跟踪算法。如图8-25所示，比如我们要让四轴飞行器跟踪一名地面上行走的女孩，该算法分为4个阶段。

(1) 初始化阶段：提取目标特征

所谓初始化，也就是在视频的第1帧，人工地指定跟踪目标，以后的画面帧就靠计算机视觉来自动跟踪了。具体来说，在视频的第1帧画面，人工用鼠标框出一个跟踪目标，这有点像我们给这个女孩拍了一张特写照片。然后算法自动提取出目标的特征，如女孩特写照片的颜色

直方图。

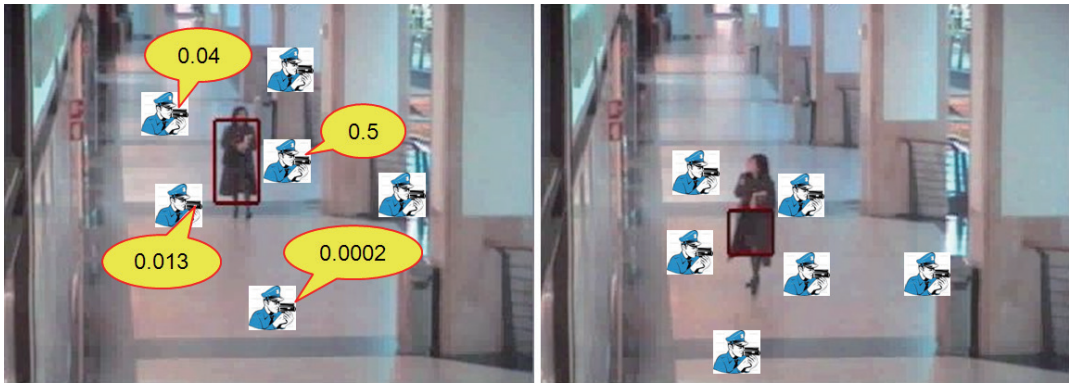


图 8-25 基于粒子滤波的目标跟踪



提示：除了人工指定目标，你还可以使用目标检测算法来自动发现目标。有简单的也有复杂的方法。最简单的是直接进行背景减除，稍微复杂点的有背景建模（基于像素的高斯混合模型、基于区域的 LBP 纹理模式、基于帧的背景子空间提取等）、光流场估计等。

(2) 搜索阶段

好了，画面到了第 2 帧，此时就要完全依靠计算机来跟踪了。计算机已经掌握了目标的特征，于是派出很多个“警察”，在这帧画面上检测目标对象，这里的“警察”是粒子（Particle）的形象比喻。怎么个派法呢？可在上一帧得到的目标（即“女孩”）附近按照高斯分布来放，可以理解成，靠近目标的地方多放，远离目标的地方少放。“警察们”派出去后，每个“警察”怎么搜索目标呢？就是根据目标特征（颜色直方图），形象点说就是根据我们提供给每个“警察”一张“女孩”的特写照片。每个“警察”在他所处的位置（某个图像像素坐标）也拍了一张照片，提取出照片的特征（颜色直方图），并计算该照片与“女孩”特写照片的相似度，得到一个数值。每个“警察”各自算完相似度后，将所有值相加得到一个总和，以此对每个相似度做一个归一化，使得所有“警察”得到的相似度加起来等于 1。这个相似度就近似表示了“警察”找到这个“女孩”的概率。

(3) 决策阶段

我们派出去的一个个精干的警察向我们发回报告，“一号警察所在处与目标的相似度是 0.5”，“二号警察所在处与目标的相似度是 0.04”，“三号警察所在处与目标的相似度是 0.000 2”，“四号警察所在处与目标的相似度是 0.013”……那么，目标究竟最可能在哪里呢？我们可以选相似度最大的“警察”所在处的图像像素坐标，比如一号“警察”所在处的像素坐标；也可以对所有地点做加权平均，如设第 i 号“警察”所在处的图像像素坐标是 (x_i, y_i) ，他报告的相似度是 w_i ，于是目标最可能的像素坐标为：
$$x = \sum_i x_i \times w_i, \quad y = \sum_i y_i \times w_i$$
。这里 \sum_i 是求和的意思，即对所有的粒子（“警察”）的坐标进行加权求和。

(4) 重采样阶段

既然我们是在做目标跟踪，一般说来，目标（即那个“女孩”）是跑来跑去乱动的。在新的帧图像里，目标可能在哪里呢（我爱的人已经飞走了）？我们在新画面里接着派“警察”搜索吧！仍然在上一帧得到的目标附近按照高斯分布来放，即靠近目标的地方多放，远离目标的地方少放。举例来说，前一帧打探的结果，一号“警察”处的相似度最高，三号“警察”处的相似度最低，于是我们要重新分布警力，正所谓好钢用在刀刃上，我们在相似度最高的“警察”附近放更多的“警察”，在相似度最低的“警察”附近少放“警察”。这就是**重要性重采样（SIR, Sampling Importance Resampling）**，即根据重要性重新放粒子，而重新采样可克服粒子退化问题。



提示：重要性采样实际上是基于所谓的**蒙特卡洛（Monte Carlo）**模拟方法，其利用随机采样对一个目标函数做近似。例如，求一个稀奇古怪的形状S的面积，如果我们没有一个解析的表达方法（如正方形的面积函数可解析表达为边长的平方），那么怎么做呢？蒙特卡洛法告诉我们，你只要均匀地在一个包裹了这个形状S的正方形内随机撒点，并统计点在形状S内的个数，那么当你撒的点足够多的时候，形状S的面积就可近似=（在形状S内的点个数/正方形内的总的点个数）×正方形的面积。

(2) → (3) → (4) → (2) 如此反复循环，即通过重采样的方法撒粒子、粒子扩散、状态观察、目标预测，直到完成了目标的动态跟踪。基于粒子滤波的目标跟踪是不是很简单？



扩展：给定**线性动态模型**，其包含两个方程：

$$\text{状态方程：} \mathbf{x}_i = \mathbf{A}_{i,i-1} \mathbf{x}_{i-1} + \mathbf{n}_i$$

$$\text{观测方程：} \mathbf{y}_i = \mathbf{C}_i \mathbf{x}_i + \mathbf{v}_i$$

其中 \mathbf{x}_i 是 i 时刻的状态向量， \mathbf{y}_i 是 i 时刻的观测向量（来自实测图像）， $\mathbf{A}_{i,i-1}$ 是状态转移矩阵， \mathbf{C}_i 是观测矩阵， \mathbf{n}_i 和 \mathbf{v}_i 分别是系统噪声和观测噪声。在给定观测值 $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ 之后，我们要对系统状态 \mathbf{x}_i ，比如 \mathbf{x}_n ，进行某种最佳估计。如果我们能通过某种方式获得后验概率密度 $p(\mathbf{X}_n | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n)$ ，那么估计起来将不会太困难。实际上，在线性动态模型假设下， $p(\mathbf{X}_n | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n)$ 是高斯的，因而 \mathbf{x}_n 的估计就可用**Kalman（卡尔曼）滤波**这种有效的递推算法来实现，实际上**Kalman滤波正是线性高斯模型下的最优状态估计算法**。然而，如果是如下的**非线性动态模型**，如：

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\mathbf{x}_{i-1}) + \mathbf{n}_i$$

其中 $\mathbf{f}_{i,i-1}()$ 为某个非线性函数。这时 $p(\mathbf{X}_n | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n)$ 显然不再是高斯的，这将使得后验概率 $p(\mathbf{X}_n | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n)$ 的形状很不规则，估计 \mathbf{x}_n 变得困难，**Kalman滤波将不适用**。

而**粒子滤波**是求解后验概率的一种算法，通过非参数化的蒙特卡洛模拟方法来实现递推贝叶斯估计。核心思想是以互相独立的样本集合 $\{\mathbf{u}_i\}_i$ 和与样本 \mathbf{u}_i 相对应的**重要性权重（Importance Weight）** w_i 来实现对后验概率密度的近似表示，并随着时间的变化对后验概率密度进行递推更新。

具体地，通常目标状态的后验概率无法直接得到，因此可根据贝叶斯重要性采样定理给出一种近似计算方法：目标状态后验分布用一系列离散的粒子来近似表示。粒子滤波用粒子集来表示概率，通过从后验概率中抽取的随机状态粒子来表达其分布，核心算法是**序贯重要性采样 (Sequential Importance Sampling, SIS)**。该算法的主要思想是利用一系列随机样本（即粒子）的加权和表示所求的后验概率密度，得到状态的估计值，当样本点数增至无穷大时，结果接近于最优贝叶斯估计。

8.6.2 基于 Mean Shift (均值漂移) 的目标跟踪算法

怎么样，现在是不是觉得跟踪有点意思？下面就趁热打铁，再介绍一种称为 **Mean Shift (均值漂移)** 的目标跟踪算法。与粒子滤波有些不同，这次我们引入搜索区域的概念。

首先，我们也是要告诉计算机需要跟踪哪一个物体，在视频的第 1 帧我们用小窗口框出需要跟踪的运动目标，如一辆红色小轿车（见图 8-27 左边），然后提取这辆红色小轿车的统计特征（比如颜色直方图）。

从第 2 帧开始，计算机就开始自动搜索了。它把搜索区域（图 8-26 中的蓝色圆框）中每个像素点处与红色小轿车的颜色直方图进行比较，为搜索区域生成了一张概率密度图。概率密度越大的地点表示包含目标的可能性越大，反之越小。

然后我们让搜索区域的中心沿概率密度增加最大的方向移动（也即向着密度中心靠拢），直至移动到目标的真实位置。如图 8-26 左边所示，我们将搜索窗口中心的黄色移动箭头称为**均值漂移向量**。随着搜索的逼近，向量的大小变得越来越小，也就是黄色箭头越来越短，直到长度为零停止移动，如图 8-26 右边所示，此时就定位到目标了！

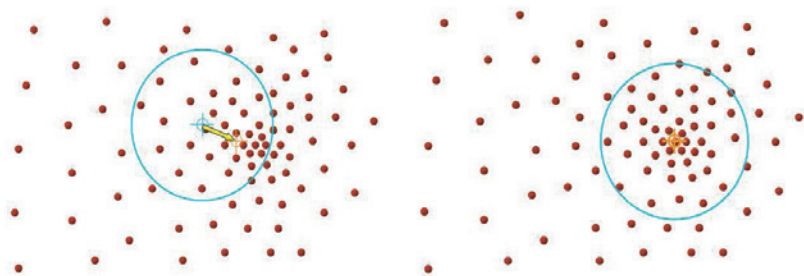


图 8-26 均值漂移向量的移动过程，直至移动到目标中心点（图片来源：tistory）

整个跟踪过程是不是更简单？这里所说的均值，几何意义上可看作是概率密度空间的重心。均值漂移的过程实际上是要在概率密度空间中寻找最密集的重心位置，直到（蓝色）搜索窗口的中心与概率密度空间的重心重合。此外，可通过定义核函数，让各个样本对漂移向量的贡献是非均匀的，也即对不同样本点设置不同的权重系数，使它们对漂移向量有不同的影响，比如离得越近则影响越大，这样可保证 Mean-Shift 获得好的性能和收敛速度。

以上就是 Mean Shift 算法过程，视频跟踪结果如图 8-27 右边所示。这里我们的搜索区域为一个红色的小方框（框中的数值为算法的某个中间结果，请直接忽略掉它）。

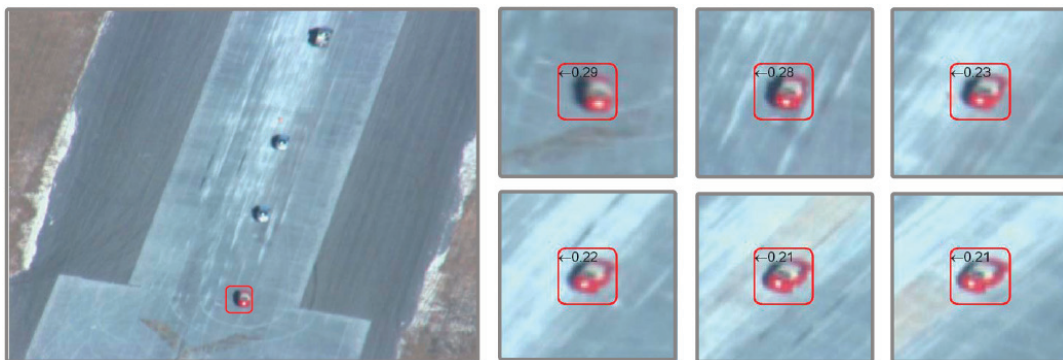


图 8-27 Mean Shift 对一辆红色小车的跟踪结果



提示：Mean Shift 是基于**核密度估计**的**非参数**特征空间分析方法，自适应步长迭代寻找概率密度分布的局部极值点。所谓**参数估计**就是密度函数的形式可简单地假设出来，例如假定是正态概率密度函数，则只需估计均值和方差这几个参数即可。但在实际应用中，数据的模型几乎都是未知的，这时**非参数估计**就大有可为。非参数方法无须预先假设密度函数的结构形式，可由落入某一连续点邻域中的若干样本点估计出其密度函数值，可实现对任意分布的密度估计。典型的无参数密度估计方法有最近邻域法、直方图法及核密度估计法。与直方图法相比，**核密度估计法是一种平滑的非参数估计方法，它增加了一个平滑数据的核函数。**

具体地，给定 d 维特征空间中 n 个样本 $\mathbf{x}_i \in \mathbf{R}^d$, $i=1, \dots, n$ ，可以得到空间中任意位置 \mathbf{x} 的核概率密度估计 $\hat{f}(\mathbf{x})$ ：

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K_H(\mathbf{x} - \mathbf{x}_i)$$

式中 $K_H(\mathbf{x})$ 为核函数。密度估计的过程可理解为：某采样点概率密度函数的估计值的取值是以每个采样点为中心的局部函数的平均值。为了分析数据集合中密度最大数据的分布位置，可对上式求梯度并从中提取出均值漂移向量：

$$m(\mathbf{x}) - \mathbf{x} = \frac{\sum_{i=1}^n K(\mathbf{x} - \mathbf{x}_i) \mathbf{x}_i}{\sum_{i=1}^n K(\mathbf{x} - \mathbf{x}_i)} - \mathbf{x}$$

其中核函数 K 用于对每个采样点进行加权：离中心点越近，则对估计中心点周围的统计特性越重要。在每次迭代过程中，将新的 $m(\mathbf{x})$ 赋给 \mathbf{x} ，然后重新计算均值漂移向量，直到 $m(\mathbf{x}) - \mathbf{x} \rightarrow 0$ 。Mean Shift 算法不断地将位置移向数据均值的地方，同时因为梯度指向密度变化最大的方向，故 **Mean Shift 向量总是指向密度增大的方向**。此外，移动的步长也与该点的概率密度有关，**在密度大的地方（如包含感兴趣特征的数据区域），Mean Shift 算法会使得移动的步长变小、漂移得慢，反之就变大**。因此，Mean Shift 算法是一个变步长的梯度上升算法，或称自适应梯度上升算法。

在将 Mean Shift 算法用于目标跟踪时，一般利用核函数加权颜色直方图描述目标，并用 Bhattacharyya 系数度量目标与候选目标之间的相似性，接着将搜寻相似度量的局部极值问题转化为核密度估计问题，然后通过 Mean Shift 迭代来寻找其局部最大值。

如果采用 OpenCV 库(请见第 6 章 6.11 节“OpenCV 与 OpenGL: 视觉计算入门的两大利器”), Mean Shift 算法的实现非常简单, 调用 meanShift() 函数即可。此外, 你还可以在 OpenCV 中调用 camShift() 函数来实现 CamShift 算法, 它是 Mean Shift 的一个改进, 称为连续自适应的 MeanShift 算法 (Continuously Adaptive Mean Shift), 其优点在于当目标的大小发生改变的时候, 此算法可以自适应调整目标区域继续跟踪。但需要指出的是, 目前 Mean Shift 算法经过学术界的不断改进, 也有这种自适应大小的能力了, 且成为了最常用的跟踪算法, 只不过 OpenCV 库中的 meanShift() 函数还没有得到相应升级而已。

第9章

3D打印之不远的将来

《荀子·非相》有云：“以近知远，以一知万，以微知明。”又云：“欲观千岁，则数今日；欲知亿万，则审一二。”相信通过前面章节的详细介绍，读者们对3D打印不远的将来已能做一些展望，正所谓“以近知远”。本章将与大家一起对3D打印的前景做具体的探讨。由于3D打印无疑将渗入到我们未来生活工作的方方面面，如果要进行全面阐述，恐怕是本书的有限篇幅所不能及的，所以我们选择了一些有代表性的应用前景，以达到“以一知万”的目的。同时还希望读者能够“欲知亿万，则审一二”，在某几个点上做深度思考，或许，你人生中的某些好机会将由此降临。

对于我国来说，3D打印带来了“中国制造”向“中国智造”转变的历史性机遇。如果找对节拍、与高科技时代共舞，“中国智造”对“全球第三次工业革命”无疑也将会产生巨大的推动作用。当然，目前这还只是一个愿景，哪怕它已并不遥远，借用孙中山先生的一句名言：“革命尚未成功，诸君仍需努力”。所以，请各位读者一起加入3D打印的热潮，发挥火热的创造热情，一起来实践“全球第三次工业革命”吧！

9.1 3D打印的未来：由创客们决定

在2012年2月7日举行的第二届“白宫科学展”（White House Science Fair）上，一群创客受邀前往白宫国宴厅，演示了来自民间的创造。如图9-1所示，14岁的创客乔伊·哈迪（Joey Hudy）在美国总统奥巴马的帮助下，发射了他自己发明的“顶级棉花糖大炮”。



图9-1 14岁的创客给美国总统奥巴马演示自己发明的“顶级棉花糖大炮”（图片来源：nationalpost）

2012年，在荷兰首都阿姆斯特丹，一场聚焦3D打印技术最新成果的现代艺术展在这里拉开帷幕。展会上精彩展示了利用各种新型材料、借助3D打印技术制作出来的时装、饰品、玩具等。其成熟的技术及炫目的效果，让人叹为观止。

9.1.1 几乎为零的设计和制造门槛

通过上面两个例子可以看到，民间的创意是无穷的，一位伟人曾说过：“人民，只有人民，才是创造世界历史的动力”。3D打印和3D智能数字化的珠联璧合，使得产品的设计和制造的门槛变得几乎为零，使得广大普通用户得以轻松“组团”成为创客，创造新工业革命的奇迹。

- 3D智能数字化的不断发展，使得产品设计越来越简单、越来越傻瓜化、越来越“所想即所得”，任何人只要有想法，哪怕没有任何设计基础，也能将自己的创意变成3D图纸。
- 同时，伴随着创客开源共享精神的发扬光大，各种各样产品的机械图纸、电路设计，以及控制源代码都能在网上轻易获得，使得你“不必重新发明轮子”。在开源社区的帮助下，你还可以和其他创客并肩作战，甚至找到创业的良师益友。
- 3D打印技术则直接让你的图纸变成了产品。不再需要购置昂贵且难以操作的大工厂机器，再也不用学徒8年去掌握各种车、削、铣、锉等加工工艺。你只需买一台小小的3D打印机放在家里，泡上一杯香浓的咖啡，电脑里点上一支悠扬的小曲，伴随着你轻盈的舞步，你的产品就直接在家里制造出来。而且，可选用的打印材料种类将越来越多，高强度的塑料已经出现，让你的产品不再只是个玩具。

当产品制造出来之后，产品标签既不是“Made in China”（中国制造），也不是“Proudly Made in USA”（美国荣誉出品），而是“Designed by Yourself”（自己智造）！当“个人智造”、“家庭智造”、“网络社区智造”所引发的第三次工业革命浪潮汹涌到来之际，任何大工厂、大公司、超级大国的巨型制造机器都将被时代大潮无情地拍在沙滩上，樯櫓灰飞烟灭。甚至可以说，设计和制造将迎来“全民皆兵”——创客们的时代，人类的创造力总当量一下子拉向了无尽的天空。

当年，福特需要花费巨资来建造他那间位于胭脂河旁的庞大汽车工厂，而他的现代同行——Local Motors的创客们（详见第1章1.5.1节“以小博大：创客挑战巨头公司”）除了一台随身携带的笔记本电脑和发明的渴望之外，一切几乎可以从零开始。

9.1.2 创客成就3D打印

本书的前文已提到，正是由于创客，才将原本极其昂贵的3D打印机从“高端大气上档次”的大工厂请到了寻常个人家。因此，创客是名副其实的个人3D打印机的创造者。创客们并没有停下脚步，而是将3D打印技术推向了一个又一个高度：从最新的生物打印进展，到食品打印机的出现，到第一把3D打印枪支的横空出世，再到3D打印房屋，将荒漠的沙子3D打印成玻璃等，这些技术的背后都是个人的身影，大公司出人意料地总是落后一步。正如有句话所说的“高手在民间”，创客的全民创造热情一旦激发出来，将远远抛下那些看似装备精良但实已因循守旧的大公司和大财团。

创客们所推动的软硬件开源已成为快速发展的科技潮流。由5名创客开发的Arduino开源硬件不仅催生了个人3D打印机，目前也已被大公司Google选作Android开源配件的开发工具。

而另外 3 名创客只花了 1 个月的时间，便做出了原型的 Square 移动支付装置，该装置在 2 年内就已处理了 20 亿美金以上的交易。以前，我国国内由于全民创造氛围的缺失造成了大量的恶性竞争和山寨的横行；如今，通过创客精神的再造，并与原有的高效率生产链相结合，已经逐步让国内的几家开源硬件公司在国际上打出了自己的品牌。确实！创客与开源出现之后，没有必要再去盗版和山寨，只需找准细分市场，每一个细小的改进和特色都可以让创客们获得相应的丰厚回报。

民众不仅创造力是无穷的，而且一旦他们成为个体制造者、不再被动地为别人打工时，对经济利益的追求也将发挥无穷的主观能动性。因此，创客们将以极大的热情推动 3D 打印应用往前发展，大大拓宽其领域范围，“无孔不入”地渗透到经济生活的方方面面。实际上，人们已经开始摩拳擦掌准备在 3D 打印机的协助下生产自己的产品了。如图 9-2 所示，它可能是一双鞋子，比如有人将直接把 3D 打印机摆在自己的鞋店里，来一位顾客就按照对方的尺码和脚型现打现卖，当然，顾客也完全可以把自己的名字或爱侣的名字打在鞋上。它也可能是件个人小饰品，在顾客选定颜色和材质后由 3D 打印机直接制成。此外还有更多的应用，比如蛋糕定制打印、3D 人像打印、玩具打印等。



图 9-2 直接将 3D 打印机摆在鞋店，按照顾客的脚步和尺码现打现卖（图片来源：3D System）

同时，创客们也在不断地升级自己的装备、“鸟枪换炮”，以便跟大公司展开有力的竞争。从家用电器到山地车，高强度塑料、碳纤维正在取代钢铁和铝，成为制造这些产品的主要材料。纳米技术让产品的功能变强，如止血的绷带、效率更高的引擎和更易清洁的瓷器。这些都将变成创客们手中的利器，一旦成功驾驭，创客们会进一步将 3D 打印发扬光大，走进日常生活的方方面面，做出与大工厂相匹敌的、极具个性化特色但价格却低廉很多的且同样“高端大气上档次”的产品出来。

大企业应当未雨绸缪，因为 3D 打印这种即将成熟的技术将会使中小企业甚至个体企业家变得更具有竞争力。产品创新将会更加容易，成本更低。供应链的地理格局也将转变。比如，在

沙漠中央工作的人发现自己缺少某件工具，他不必再让人从离他最近的城市买来。这时他摇身一变成为一名创客，只要简单地下载一份工具设计图纸，然后 3D 打印出来即可。关于创客与大企业未来的格局分布，请阅读第 1 章 1.5.2 节“聚沙成塔：改变工业社会的组成结构”。

9.2 手机应用FabApp、App Store与智能云网

App 是英文 Application（应用程序）的简称，由于 iPhone 智能手机以及移动互联网的流行，App 一般特指智能手机的第三方应用程序。App Store（应用程序商店）是第三方应用程序的开发者与用户之间的连接桥梁。App Store 平台上大部分应用价格低于 10 美元。用户购买应用程序所支付的费用由苹果与开发者 3:7 分成。截至 2013 年 1 月 7 日 App Store 上共有 103 万个 App，总下载量已经突破 400 亿次。App Store 平台形成了一个良性的“生态环境”，即“优质应用→优质用户→优质应用”这一良性循环。

3D 打印的出现激发了人们随时随地想制造东西的念头，比如今天心情不错，想给自己造一块手表。然而手表是需要多年的经验和技巧才能设计成功的，怎么办？在你的手机上下载一个“造手表”的 FabApp（Fab 是 Fabrication 的缩写，制造）！一个 FabApp 就像一个 iPhone 应用程序一样，如图 9-3 所示，你可以自己定义手表的颜色和表带的样式，以及自定义表带上的刻字和签名，还可上传几张你手腕的照片以确保完美的贴合。然后下单，造表厂就会根据你的定制进行 3D 打印并立即发货。因为无须库存，所以一个 FabApp 的价格也非常低廉，比如“造手表”的 FabApp 只需要 10 元。

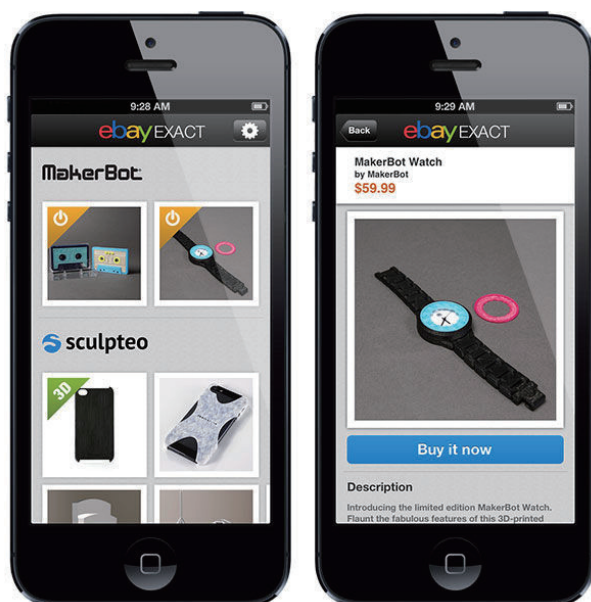


图 9-3 eBay 提供 FabApp 手机应用用于定制一块手表

注意上面有一个细节：你可上传几张手腕的照片，然后服务器端自动根据这些照片智能地计算出你手腕的各个尺寸。然而，普通的造表厂一般不具备这种高级的 3D 智能化技术。但他们

也不必自己开发，正如在第4章4.5节“智能云网：云端智能服务和云制造”介绍的那样，他们只需购买云端智能计算服务API即可。像iPhone应用程序一样，FabApp将产生新的经济。定制化打印应用程序将在那些有限而复杂的市场中发挥优势，而这些利基市场可能太小，不足以吸引那些大型制造企业，但却足够为小型企业和个人提供机会。

再举一个青年男女可能会喜欢的例子。比如你在街上遇见一个心仪的女孩/男孩，用iPhone拍下她/他的照片，然后花10元钱购买一个“3D Girl”或“3D Boy”的FabApp，云端服务器自动根据你上传的照片将她/他的3D照片重建成3D模型，然后按照真人1:1打印出来寄送到你的家里。当然，这将毫无疑问地涉及肖像权问题以及法律上可能的制裁，为降低风险，请仔细阅读第9.9节“枪支打印‘让子弹飞’、版权与社会伦理”。

9.3 不再仅仅是看着粗糙的FDM

目前3D打印最容易引起反对者非议的地方是：桌面级3D打印机一般都采用FDM（熔融沉积成型）工艺，这种工艺精度不高，且采用的ABS和PLA塑料耗材强度也不高，无法真正用于产品制造。那么，究竟是什么限制了当今的3D打印技术，阻碍这项可能彻底改变人类工业面貌的技术的发展？这就是专利。但是在2014年1月，一个阻碍创客们获得高端3D打印技术的专利已经到期了。该专利覆盖了一项关键的技术：激光烧结，它其实是成本很低的一种打印技术。激光烧结在任何轴向上都拥有高分辨率，材料可以使用金属和陶瓷，可以打印出完美的最终以销售的产品。



3D打印技术的研发始于20世纪80年代中后期，相关的基本专利也都于20世纪90年代开始申请，但目前这些专利大部分已经或即将因有效期届满而失效。

目前，工业上使用的高端激光烧结3D打印机是普通个人和艺术家设计师负担不起的，因为每次可能要花费上万美元。而随着激光金属烧结技术专利的到期，我们将会看到相关3D打印机设备价格的迅速下跌。举个近在咫尺的例子。正是因为前几年FDM（熔融沉积成型）专利的到期，才引发了个人3D打印机的诞生，并由此衍生出上百种开源FDM打印机，比如著名的个人3D打印机MakerBot。相应地，在FDM相关专利到期后的数年间，FDM打印机的价格从14 000美元垂直下降到300美元。这直接催生了一个3D打印爱好者市场，培育了一大批玩家，在家里打印好玩的微缩模型。

在专利技术壁垒被打破以后，这样的事情也会发生在高端激光3D打印机市场，价格的剧跌将加剧3D激光打印机市场的竞争，专利公开以后，所有技术细节将被开源化。今后，大部分廉价3D打印机会从哪儿来呢？当然是中国。2012年，中华人民共和国工业和信息化部投资2亿元人民币，启动了10个3D打印研发中心。

今后，随着激光金属烧结专利的到期，设计师可以在几个小时内把设计变成实物，甚至直接对外销售——比如Google Glass的支架，本来设计师需要通过专业3D打印服务公司才能得到成品，到那时直接在家用激光烧结打印机制造出来即可。

目前已开始有创客跃跃欲试。比如针对专利同样到期的 SLA（光固化立体造型）技术，已推出廉价的 3D 打印机 Formlabs Form 1。SLA 工艺的精度比 FDM 要高得多，而一台 Formlabs 只要 3 300 美金，就可以让用户自己打印高质量的树脂模型，如图 9-4 所示。Formlabs 也遇到过专利问题，以至于曾被专利的持有者 3D Systems 告上法庭，但后者的专利已经到期了，因此最终也只好不了了之。



图 9-4 Formlabs 的桌面级 3D 光固化打印机 Form 1

这一切都意味着，大规模产品定制的工业生产时代即将到来。

9.4 生物医疗打印：越来越近的科幻

3D 打印尤其适合于医学领域。比如在修复性医学领域，个性化定制的需求十分明显。用于治疗个体的产品，基本上都是定制化的，不存在标准的量化生产。而 3D 打印技术的引入，降低了定制化生产的成本。随着全球老龄化程度的加剧，修复性医学中的结构性器官移植将会持续增长，特别是牙科领域。值得一提的是，跟军事领域一样，医学领域的产品一般不考虑性价比，不太计较成本，是个典型的高利润行业。

现阶段，3D 打印技术在医学领域的主要应用在于如下几点。

- 修复性医学中的人体移植器官制造，假牙、骨骼、假肢等，如利用 3D 激光成型技术制作的钛合金移植颞骨。
- 辅助治疗中使用的医疗装置，如牙齿矫正器、助听器等。
- 手术和其他治疗过程中使用的辅助装置，如在脊椎手术中，用于固定静脉的器械装置。

以骨骼为例，加拿大有一所大学目前正在研发“骨骼打印机”。这种打印机可以使用人造骨粉作为耗材，把这些骨粉转变成精密的骨骼组织，如图 9-5 左边所示。这种人造骨骼表面布满孔隙，它们像海绵一样可以将周边的骨头吸引进来（即具有骨导性），使真骨与假骨之间结成牢固

的一体，患者骨骼能尽快康复。

当然，目前使用最多的还是钛合金骨骼。有患者担心，3D 打印钛合金骨骼价格肯定不菲。而实际上，用 3D 打印技术生产出来的产品与传统工艺产品相比，成本不会增加，甚至有可能降低，其价格不高于，甚至低于传统产品价格。而且，除了钛金属，目前已研发出了一种可供手术植入的新型塑料 OsteoFab，克服了金属会干扰 X 射线而无法用于高精度核磁成像手术的缺点，植入物利用 SLS 工艺进行 3D 打印成型，强度和韧度都很高。

再举一个 3D 打印人脸的例子。英国一名叫 Eric Moger 的 60 岁男子因为罹患一个网球大的肿瘤，手术后被切除了几乎整张左脸，包括他的眼睛、颧骨、颌骨，留下一个大洞。不过现在，感谢 3D 打印技术，Moger 又拥有了一张完整的脸，如图 9-5 右边所示。

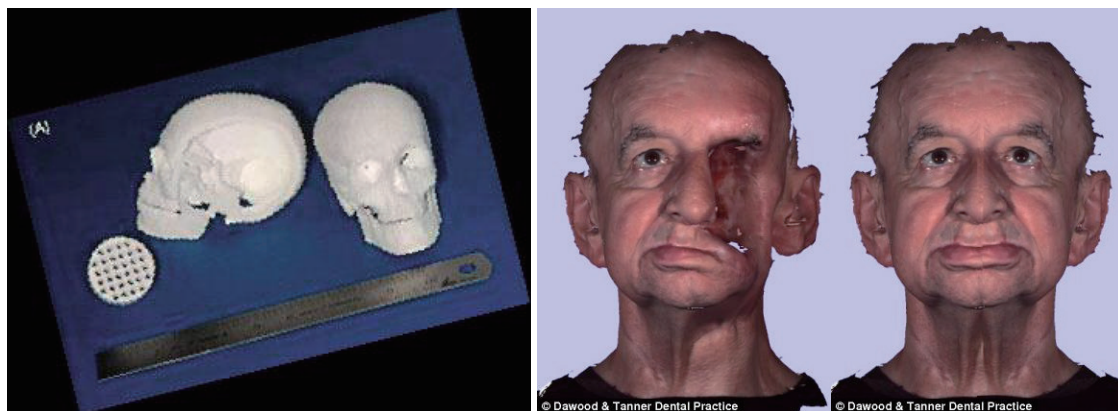


图 9-5 3D 打印用于修复性医学。左：3D 打印的人造骨骼组织；右：用尼龙打印的左脸
(图片来源：Dawood & Tanner Dental Practice)

医生首先利用 CT 扫描和面部扫描来设计 Moger 想要的脸型。接下来，利用 3D 碾磨技术将钛金属加工成一个支架并移植进左半脸。然后，使用计算机软件镜像复制了 Moger 的右脸，并用有韧性的尼龙作为材料 3D 打印他的左脸。医生说希望以后能进一步采用硅胶。

此外，3D 生物打印还能制造功能性器官，这使得应用范围扩展到无限可能。具体来说，可利用干细胞为材料，按 3D 成型技术进行制造。一旦细胞正确着位，便可以生长成器官，“打印”的新生组织会形成自给的血管和内部结构。在打印时，可将这些细胞先放入一个个球体中形成基本的打印单元，每层用生物纸（由另一个喷头所打印，主要成分是水凝胶：Hydrogel）作为细胞生长的支架，接着一个个球体被打放置生物纸上，如此层叠打印更多的生物纸和放置一个个球体。打印完成后球体会慢慢融合成为坚实的组织，并自动进行重新排列。最后，支撑用的生物纸将被溶解或移除掉。在 3D 生物打印领域领军的 Organovo 公司，已经成功打印出了心肌组织、肺、动静脉血管等。

这个技术的形成，有一个基础性的原理。科学家曾做过实验，将人类动脉血管切成一节节的环状结构，然后把这些环状结构套在一根线上，大概在 72 小时后发现，这些切开的血管又融合到了一起，形成了一根新血管。这个实验告诉我们：在体外，如果把不同的细胞在空间上按照人类组织器官细胞的排列规则放在一起，这些细胞会很快发生迁移、扩散、自组织，重新组

成一个器官。这听起来似乎难以置信，但实际上经过几十亿年的进化，各种类型的细胞都已自己知道了该如何生长成为复杂器官。

那么，3D 生物打印机具体是如何制造血管的呢？如图 9-6 所示，研究人员将一排排水凝胶平行放入培养皿的水槽里，在水槽中打印颗粒状细胞圆柱体。至少有一个水凝胶圆柱体被打印在细胞的中间位置，用于制作静脉内的小孔，血液可以通过这个小孔流进 / 流出。

通过 3D 生物技术，我们可以利用病人自己的细胞来制造所缺损的器官。由于采用的是患者自身的细胞，所以基本没有排异反应的风险。比如用自己的皮肤细胞打印一块皮肤，用于烧伤后的皮肤移植。相似地，还可用自己的脂肪细胞进行隆胸，相比于目前的硅胶隆胸，几乎没有副作用。

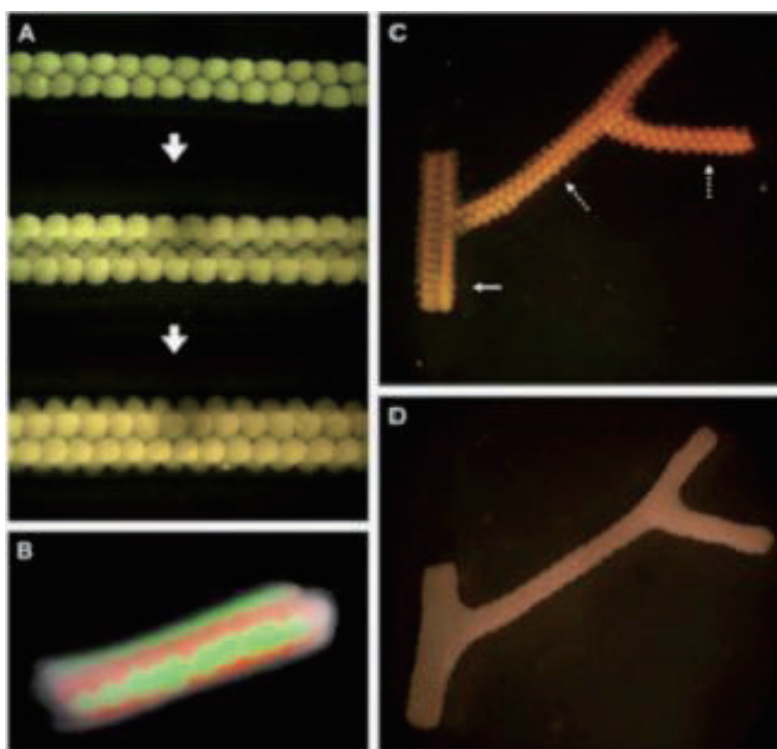


图 9-6 3D 生物打印机制造血管（图片来源：Organovo）

3D 生物打印机制造的人工内脏，比如人造肝脏、人造肾脏，目前还不能用于真正的体内移植，但完全可以用于药物研发领域的药物筛选。用 3D 打印做出这些人工组织器官，去代替小白鼠和猴子进行药物筛选，可以大大提升筛选时间和准确率，提高新药的研发速度。

3D 生物打印的革命性之处在于：不仅可以制造跟人类功能相似的器官，更可超越人类器官原有的功能。继制造出有触觉的生物人耳（不只是装饰性的模型）之后，科学家利用 3D 打印机创造了一只耳朵，如图 9-7 所示，能够听到超过人耳听力范围的无线电频率。因此，3D 生物打印未来将逐渐从仿生演化到替代、升级，极大地推动人类自身的快速进化。



图 9-7 普林斯顿大学研究人员创造的耳朵能听到超过人耳听力范围的无线电频率

9.5 美食打印机：“吃货”的钱最好赚

不会做饭？没有关系！康奈尔大学的研究小组最近研发出一台开源的 3D 食物打印机，其型号为 Fab@Home。或许它以后有潜力成为餐馆、家庭厨房的必备厨具。Fab@Home 填充黏液状的原材料，然后注射器喷射成型。你从网上下载几千个食谱，选择其中的一个输入打印机，然后按一下按钮，不用多久就能品尝美味，如图 9-8 所示。Fab@Home Model 3 打印机目前售价仅 4 000 美元，5 年内价格能降至约几百美元。

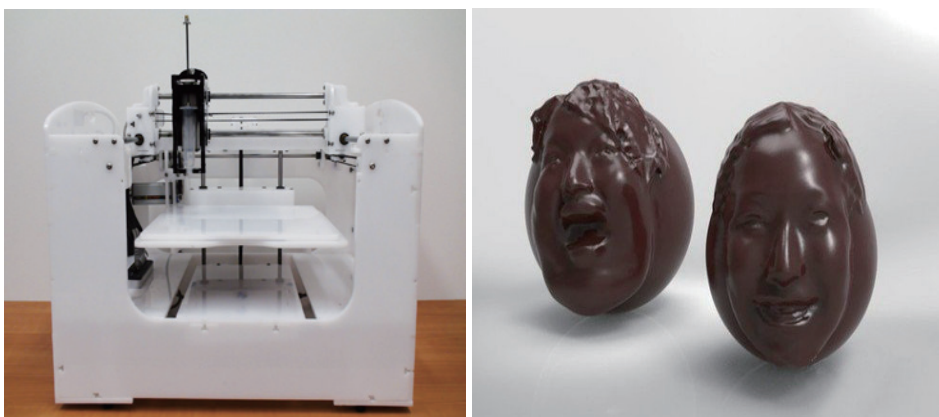


图 9-8 Fab@Home 以及其打印的巧克力食品（右边的巧克力头像不是直接 3D 打印出来的，而是先用 3D 打印机打出以医用硅酮为材料的塑模，再倒入巧克力进行浇灌）

目前，适用于这台打印机的食材仅限于“能从注射器中挤出来的东西”（如图 9-9 所示），如液体奶酪、巧克力和蛋糕面糊等，已成功地制作出曲奇饼干、奶油蛋糕等食品，有的大厨还打印出日本人爱吃的“寿司”。

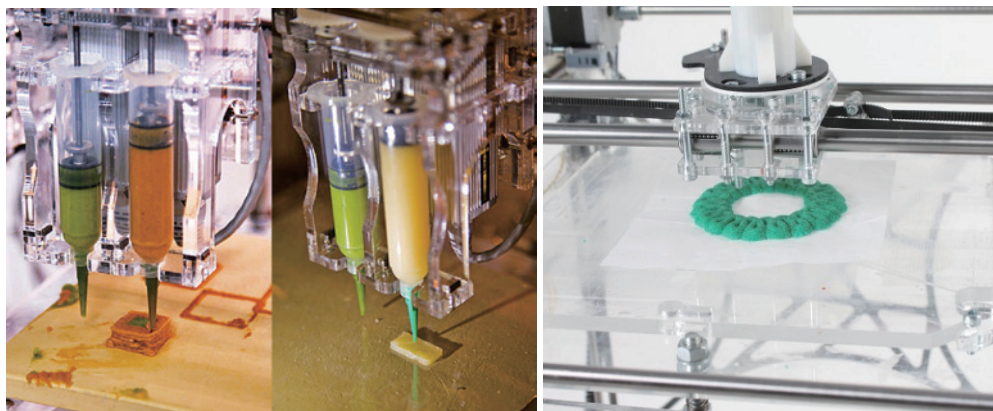
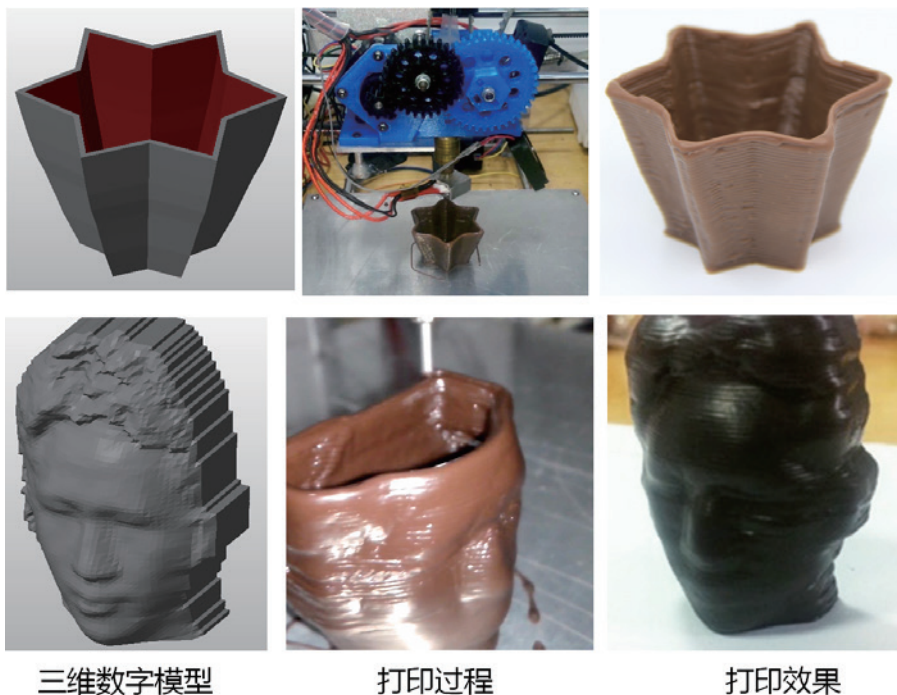


图 9-9 通过从注射器中挤出液体食材来制作食品（图片来源：howtoons）

国内也有研究人员进行食品 3D 打印的研究。最近，河海大学研发了个性化巧克力 3D 打印平台，综合实现了低成本的三维扫描、数据优化、巧克力打印功能。其打印模块硬件成本不足 3000 元，并具有较高的打印精度。如图 9-10 所示，该打印机能够实现复杂结构三维模型的直接打印，并取得非常不错的打印效果。



三维数字模型

打印过程

打印效果

图 9-10 河海大学研发的巧克力 3D 打印机及打印效果（图片来源：童晶）

一家名为 Biozoon 的德国公司正在研究可供应日常主食的 3D 打印机，如图 9-11 所示，打印的食物不仅可保持真实的形状和味道，还可以加入特定的营养素。更有价值的是，它制造出的是可在口中迅速溶化的食物，可解决老人进食困难的问题。打印机使用 48 个喷嘴，通过喷射技术，混合一些新鲜食材泥，例如芦笋、鸡肉、猪肉、豌豆、意大利面、土豆等。

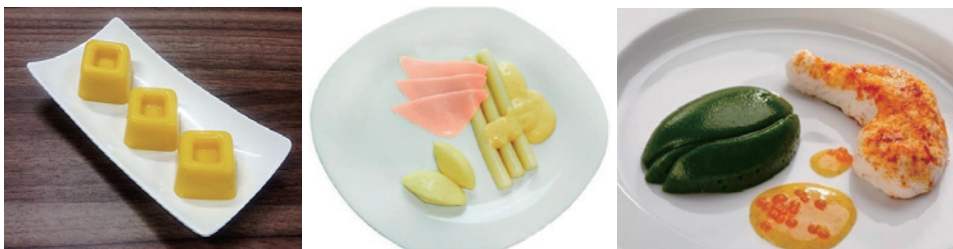


图 9-11 德国 Biozoon 公司研究的 3D 食物打印机（图片来源：Biozoon）

使用 3D 食物打印机制作食品的好处很多,如可以减少种植、烹饪和加工环节,从而避免施肥、煎炸和包装等环节对环境的影响。厨师们可以借助它来发挥创造力,制作任意形状的个性食品(如图 9-12 所示),满足挑剔食客的口味需求。医生也可以为糖尿病患者设计个性化菜肴,以适应他们在饮食上的特殊需要(如精确指定每日的糖分摄入量)。

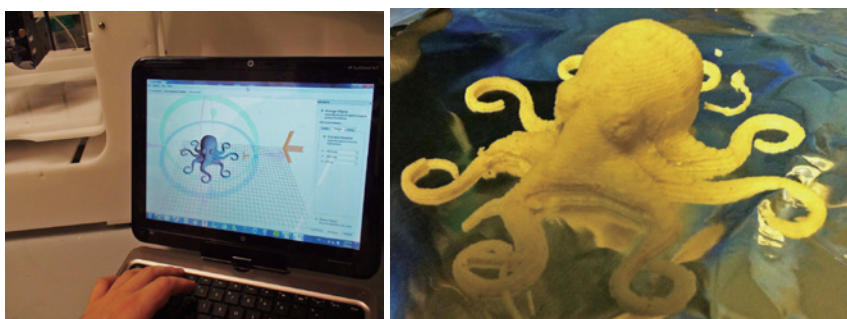


图 9-12 3D 打印的章鱼形玉米面包（图片来源：Jeffrey Lipton）

液化的原材料能很好地保存,而且可以高效利用厨房空间。很多烹饪过程中的苦恼将会消失:你再也看不到大堆的锅碗瓢盆了,同时也省下了运输各种蔬菜水果所需的昂贵费用。

下面再介绍一款名叫 CandyFab 4 000 的砂糖 3D 打印机,由 Evil Mad Scientist Laboratories (邪恶科学家实验室)制作完成,其工作原理是喷射加热过的砂糖,如图 9-13 所示。这台机器的成本不到 500 美元,精度达到 5 ~ 20 ppi (pixels per inch, 每英寸的像素数目)。

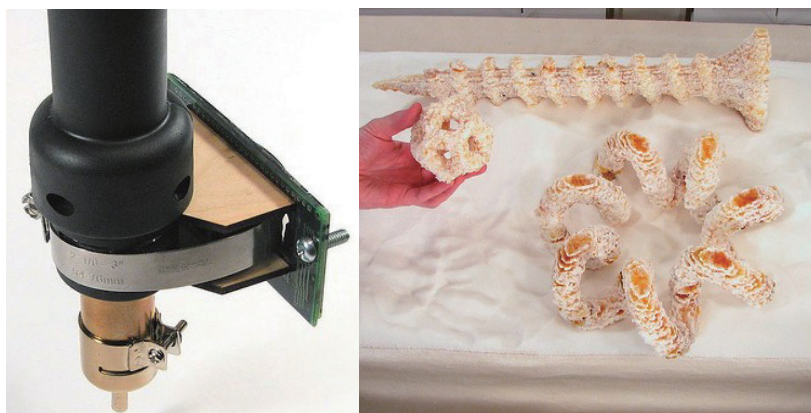


图 9-13 砂糖 3D 打印机的喷头和作品（图片来源：Evil Mad Scientist Lab）

很多北方的读者都喜欢吃街上卖的鸡蛋煎饼。有些品味雅致的读者可能会不满意摊主拿刷墙的刷子刷酱的那道工序，糊糊的一层总觉得不太美观。而有了 3D 打印，如图 9-14 所示，你可以自己在煎饼上打印精美的图案，档次和品味是不是提升上来了？当然，除了刷酱，你还可以直接用面粉糊打印出各种精美和特殊的煎饼形状。



图 9-14 在煎饼上打印精美的图案

众所周知，卡布基诺是一种既美观又美味的泡沫咖啡，制作工艺复杂，亲手做出这样的咖啡可不容易。不过目前有高手用废旧的平面绘图仪改装成了这样一款咖啡 3D 打印机。有了它，你就可以在咖啡上绘制出任何形状了，如图 9-15 所示。



图 9-15 打印卡布基诺（图片来源：ilovecoffeebook）

9.6 绿色经济：变沙漠为光影城市

土地沙漠化已成为最为严重的环境与社会经济问题，被称为地球的“癌症”，困扰了人类很多年。3D 打印技术的出现，提供了一条新的解决途径：直接将沙漠改造成道路和光影城市！

目前 3D 打印技术的一个关键就是耗材，非常昂贵。那世界上什么耗材最便宜？最便宜的莫过于沙子和阳光了。沙子和阳光真的能做 3D 打印的耗材吗？能！一个叫 Markus Kayer 的人发明了一种 3D 打印机，可以利用沙子和太阳光打印出不可思议的作品。

这种装置采用的原理为太阳能烧结（Solar Sinter），用聚焦后的普通阳光熔化沙子，制造出 3D 玻璃物体，如图 9-16 所示，而原料就是全球沙漠中无处不在的沙子和阳光。

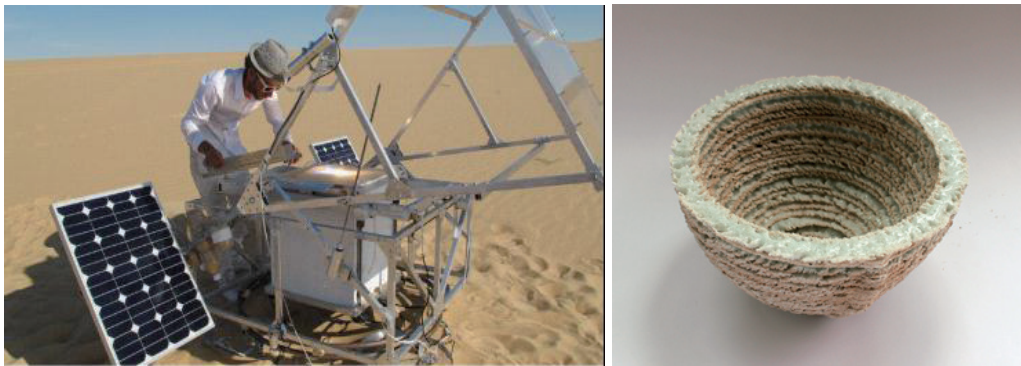


图 9-16 Markus Kayer 利用阳光和沙子打印出来的 3D 作品

而来自西班牙的建筑师则直接考虑用水、阳光和沙子来建筑桥梁，如图 9-17 所示的“鹊桥相会”场景。他们所研制的 Stone Spray（喷石）3D 打印机，可通过计算机控制，能耗很小，并可完全采用太阳能供电。在施工现场，将泥土和黏结剂混合，用喷枪在表面上喷出坚如磐石的结构。



图 9-17 喷石 3D 打印机可以打印出建筑物大小的物体（图片来源：IAAC）

9.7 打印房屋：安得广厦千万间

意大利工程师 Enrico Dini 是成功制造 3D 打印机用以打印房屋的第一人，如图 9-18 所示。他说：“我追求的结果归根结底就是让人们用可以负担得起的方式建造房子。”3D 打印房屋的概念将可能对房屋建造领域产生“革命性影响”，一旦技术成熟之后，将能解决很多地区的住房危机问题。住户住在 3D 打印制作出的房子里，无须担心“打印”出来的房子不结实，因为用于打印的材料单位面积的承重能力是标准混凝土的 3 倍。



图 9-18 Enrico Dini 所打印出来的房屋

而英国伦敦的一家建筑企业提出了 3D 打印房屋的新概念——原材料来自激光烧结的生物塑料，外观像蜘蛛网（“盘丝洞”），如图 9-19 所示。房屋以极具特色的纤维尼龙结构作为骨架，来代替实心的墙体，并用维可牢（Velcro，俗称魔术贴）尼龙搭扣或像纽扣一样的扣件将组件固定在一起。纤维结构的厚度只有 0.7mm，用石头打印是不可能的，因为沙子没有足够的结构强度和完整性。按照发明者的设计：将所有的组件制造好只需要 3 个星期的时间，装配起来则仅需 1 天的工夫。



图 9-19 利用纤维尼龙结构 3D 打印房屋（图片来源：Softkill Design）

又如国外一个名为“神奇板凳”的项目，如图 9-20 所示，内部是带有空洞的蜂窝状结构，使其重量更轻、更结实和防震。而且这种结构以后还可放置电路、电线和水管等额外的住宅基础设施，节省了空间。



图 9-20 3D 打印出的“神奇板凳”（图片来源：Agnese Sanvito）

如图 9-21 所示是上海盈创装饰打印的 10 栋别墅毛坯房，其中最大的一幢两层建筑长 10m、宽 6m、高 4m。这 10 栋房子总共只花费了 24 个小时建成，而且是整栋打印。“同样是建设两层高的建筑，传统方法要用一个多月的时间，而 3D 打印几个小时就能开发完成”，据该公司介绍，这种打印方式比传统的建筑方式节省 50% 的成本，还可以将城市所有建筑垃圾经过处理回收收到建筑中去，使建筑更加环保、节能、耐久。



图 9-21 24 小时打印出的 10 栋房屋（图片来源：上海盈创装饰）

读者可能会觉得以上的房屋都略显粗糙，那我们来介绍一个精细的。建筑师 Benjamin Dillenburger 的 3D 打印房间共计有 80 万个面，房间内部充满了复杂的装饰性设计，如图 9-22 所示。通过手工来设计是不可能的，这款作品使用的是类似细胞分裂的智能算法（参见第 4 章 4.2.1 节的细分曲面）。



图 9-22 含有 80 万个精细面的 3D 打印房间（图片来源：Benjamin Dillenburg）

9.8 混合材料制造：3D打印电路

多材料混合打印有助于制造更复杂的产品。目前领先的多材料 3D 打印机是 Objet Connex 500，最多允许同时打印 14 种塑料类材料（在此基础上可混搭出 107 种材料），可以是橡胶塑料或者更坚硬的 ABS 塑料。神奇的是，所有这些材料在一次作业任务中完成打印，在同一时间内完成融合，而不需要独立打印各个零件后再逐一组装，如图 9-23 所示。终有一日，一个完整的产品或设备可以一次打印完成，比如一次打印出一部移动电话，包括塑料外壳、金属部件、电子元件、玻璃屏幕等。



图 9-23 Objet 打印机将两种材料混合制造（图片来源：Stratasys）

下面再介绍一款 3D 打印的电池，其仅有 1mm 宽，将被应用于微型计算机。这款电池的体积比一粒沙还小（如图 9-24 所示），但面积能量密度和功率密度与手机电池相同。该电池的问世将会对微小设备领域的发展起到重大影响，比如纳米机器人以及微型医疗和通信设备，此外还可以应用在可穿戴设备上。

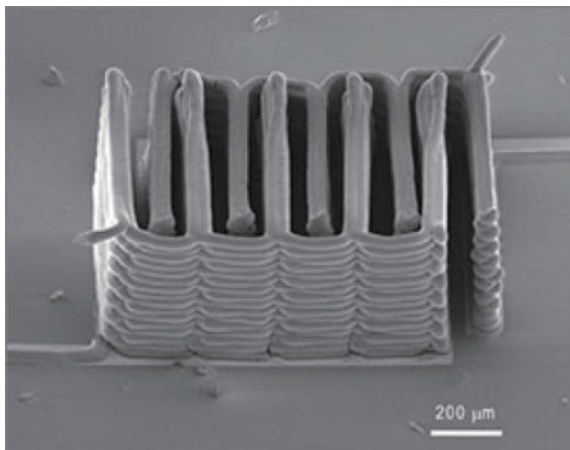


图 9-24 3D 打印的电池比一粒沙还小（图片来源：哈佛大学）

电池的打印过程有点复杂，科学家需要使用一个拥有 1mm 广角喷嘴的 3D 打印机来将两个独立的锂金属氧化物喷在梳状沉积层上，硬化之后可形成阳极和阴极。加入电解液之后，一个如沙子般细小的、充电/放电、寿命以及能量密度与普通商用电池媲美的 3D 电池就此诞生了。

有了电池，我们还需要电路板，对吧？你是否想过在家里能像打印 Word 文档一样轻松地“打印”出一张电路板，然后组装出自己喜欢的电动玩具？然而，当前的 3D 打印机大多数只能打印模型模具，还不能打印出包含电子功能在内的器件。

中国科学院的研究人员研制出室温状态下将液态金属直接印刷在纸上生成电路的技术，如图 9-24 的左边所示。传统工艺下，电子工程师若要更改电路板，需用化学药水做处理，经过刻蚀等步骤才能形成自己的设计。而新的液态金属打印方法，让漫长的设计过程变得唾手可得。打印一张 A4 纸大小的纸基电路板，目前只需要十几分钟。新方法不但可以打印平面电路，还能完成立体复杂电路及其支撑件的直接生成。

此外，国外一家名为 BotFactory 的公司开发出 Squink，能够以极低的成本在几分钟内打印出电路板，如图 9-25 右边所示。“它可以放在办公桌上，并在几分钟之内在柔性或刚性基板上打印、组装好一块电路板。” BotFactory 团队宣称。Squink 定价为 2 999 美元每台。

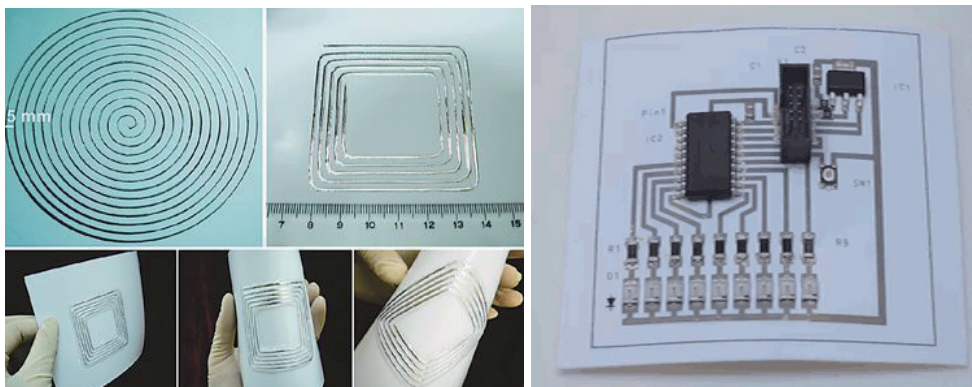


图 9-25 室温状态下，用液态金属直接在纸上印刷电子电路（图片来源：中国科学院、BotFactory）

3D 打印的便利性使得我们可以将电路结构打印在产品的外围，有助于设计产品的最佳造型，不需要再像传统的印刷电路板那样为了适应产品的外观而在形状和大小上受限。

最后我们来做个展望。**全印制电子 (Printed Electronics)** 是近年来在微电子领域出现的一种革命性的先进制造技术。它采用打印的方式将各种电子功能结构材料集成于电路板之中，除了电极以外还可打印制作晶体管、电感、电容、电阻、电池等功能组件。它具有高密度化、柔性化、集成化及环保化的优点，彻底改变当前由于采用减成法 PCB 制造技术所造成的高材料消耗、高废液量等顽症。目前，一些印制电子产品如：有机太阳能电池 (OPV, Organic Photovoltage)、射频识别电子标签 (RFID, Radio Frequency Identification)、柔性有机发光二极管显示屏 (Flexible OLED Display, 如图 9-26 所示) 等已获得市场的应用。



图 9-26 可翘曲、可折叠的柔性有机发光二极管显示屏 (图片来源 : 三星)

值得一提的是，新兴材料**石墨烯**有可能改变电子元器件的制造方式。这是有史以来最薄的材料，只有一个碳原子厚度；也是有史以来最强的材料，强度是一般结构钢的 200 倍。石墨烯的导电性与铜类似，但热传导性优于已知的所有材料。石墨烯几乎是完全透明的，但结构非常致密，即使是最小的氢原子都不能穿过它。石墨烯很适合用来制造透明触控屏幕、光板，甚至是太阳能电池。英国两位科学家因为成功地在实验中从石墨分离出石墨烯，共同获得了 2010 年诺贝尔物理学奖。目前研究人员正在研发使用石墨烯作为 3D 打印材料的技术。

9.9 枪支打印“让子弹飞”、版权与社会伦理

3D 打印所引发的第三次工业革命，将对人类生活带来深刻的影响。在大潮汹涌向前的同时，“泥沙俱下”，更不乏浑水摸鱼者。尤其是向新的社会生产关系转型的过渡阶段，人们的社会观、哲学观、价值观、道德观、伦理观都会不可避免地受到冲击和烤炙。

9.9.1 3D 打印引发社会公共安全的忧虑

美国一个名叫“分布式防御”(Defense Distributed) 的组织设计出全球第一款 3D 打印手枪，除了撞针是金属的，其他部件采用普通的塑料即可。该组织还将手枪的 3D 设计图纸放在网上供人免费下载，普通民众利用 3D 打印机在家就能轻松制造。

而在此之前不久，一位自称 HaveBlue 的网友在网站发文称自己成功打印出真枪的部分组件，并结合真枪其他部件制作成一把枪（如图 9-27 所示），还在一个农场进行了试枪。“6m 远的地方都能瞄得很准，累计射击 200 余次，枪身依然完好无损。” HaveBlue 颇为自豪地说。



图 9-27 3D 打印机所打印的枪支部件（已承受数百次实弹射击测试）（图片来源：Michael Guslick）

如果他所言非虚，他将成为世界上第一个成功使用 3D 打印枪的人。但同时也引起了公众恐慌。如果手枪能够轻轻松松打印并且成功使用，那么人人持有武器的时代就会到来。特别是在中国等大多数公民不能持枪的国家，社会问题将尤为严重。因为，它的步骤太过简单，拥有一台 3D 打印机，下载一张图纸，购买需要的材料即可。3D 打印枪一出现，立刻启发了网民的思路，手榴弹等制作简单的杀伤武器被纳入了打印的范畴。正所谓“没有枪，没有炮，3D 打印给我们造”。

为了验证打印的枪支确实可能会对社会造成危害，英国记者还专门进行了一次实验，如图 9-28 所示。他们从网上下载了手枪的 3D 图纸，然后不到 36 个小时，就用一台 3D 打印机打印出了手枪所有的塑料部件，唯一的耐磨金属部件是撞针，但这个在五金店就能买得到。最后，他们使用一些简单的工具把零部件拼装起来，不到几分钟 3D 打印的手枪就诞生了！他们把手枪拆解为 3 部分并藏在自己的衣服里，因为部件都是塑料的，并不会触发金属探测器的警报，就这样他们通过了 St Pancras 国际车站的安检。登上列车后，两名记者只用短短 30s 就把这支手枪还原了。

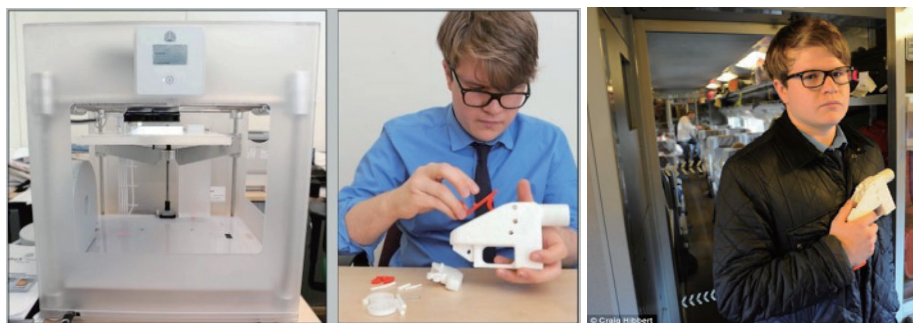


图 9-28 英国记者使用 3D 打印机打印枪支，躲过安检后在列车上快速组装（图片来源：daily mail）

这次实验为全世界的安全监测服务带来了严峻的考验，要检测出这样被分解成若干部分的塑料手枪，以现在的监测手段来看实在是太有难度了。更何况，以 3D 打印的任意成型能力，完全可以将所打印的枪支伪装成一部手机或者一个皮球。

不仅仅是武器，一些唯一性的东西，有了 3D 打印技术也不再安全，比如手铐的钥匙。前不久，德国的一名警察就通过 3D 扫描打印出了手铐的钥匙并成功开锁。当然，对于白领来说也有“福音”，公司的指纹打卡机将会形同虚设。

在澳洲，悉尼警察最近逮捕了一伙嫌犯，他们使用 3D 打印机和 CAD 设计图纸制造出了复杂的 ATM 扫描盗取装置，非法获取了超过 10 万澳币。这种装置整合度高并附带极其微型的摄像机，特别隐蔽根本看不出是另加装上去的，很难被取款人发觉。

当然，器物的强大本身并不可怕，关键要看是否是有道义的一方在掌握它，正所谓“魔高一尺，道高一丈”。日本警视厅就利用 3D 打印技术成功追捕了奥姆真理教的最后一名被告——高桥。警视厅的科学搜查研究所利用高桥的二维画像制作了脸部 3D 模型，如图 9-29 所示，没想到 3D 模型一经 ANN 电视新闻的公布报道，高桥就被抓获了。



图 9-29 日本警视厅利用 3D 打印技术成功追捕罪犯（图片来源：ANN News）

9.9.2 版权保护的难题

神奇的 3D 打印技术正在全球悄然兴起。然而，近来发生的几起因版权问题导致的叫停事件，给这一新兴行业蒙上了一层阴影。据媒体报道，一家英国游戏公司给某 3D 打印商家发出了停业“命令”，原因是该商家用 3D 打印机擅自制作了该公司的流行桌面游戏“战锤”中人物的 3D 实体模型。

大公司的固有市场正遭遇小企业的侵袭，为此它们打出的第一张牌便是法律手段，以打击那些由 3D 打印生产出来的高科技产品。华盛顿的专利和商标律师达雷尔·莫特利(Darrell Mottley)称：“我们来到了一个临界点，科技的门槛不再那么高不可攀，获取科技的成本在下降。如果你是一家成熟的制造商，当看到普通人都能够生产出替代你公司产品，你会做何感想？那意味着什么？”

由于 3D 打印摒弃了传统的注塑成型等生产环节，让生产过程变得简单直接。设计师可以自己设计，或者干脆在网上下载模型图，就能自己打印出产品。不像现在的网络数字化虚拟复制（如盗版软件、盗版音乐），3D 打印是真正“实打实”的复制，这让吃设计饭的人担忧。从网络信息

共享时代吃尽苦头的创作人刚刚摸出了赚钱途径，又被 3D 打印的到来重重打击。3D 打印成熟之后，版权问题将层出不穷。

当某款新手机（譬如，iPhone 9？）出来的时候，山寨厂商只需扫描一下，3D 打印一下，山寨机就会源源不断。特别是制造业上的巧妙设计，将毫无秘密可言。以设计为主要成本的衣服、鞋子、名牌包将会被山寨得一模一样。此外，把路易威登新出的帆布包，打印成山寨版的皮包也不是不可能，只要用户选择了合适的材料。再比如，现在用塑料 3D 打印出乐高玩具已不是问题，水平高的玩家甚至能自己修改设计，打造出属于自己的乐高。

据报道，中国发明家协会组织了一些企业到北京计算中心参观 3D 打印机，大多数人对这一技术不敏感，不觉得它物有所值，唯一产生兴趣的是一家想做山寨眼镜的企业。可见，3D 打印技术天生对于山寨有着巨大引力。

2D 时代，偷拍、艳照泛滥，以及 PS 技术移花接木，如果把它们都变成 3D 的呢？将这些 2D 照片转换成 3D 模型并不复杂，再把它们打印出来亦非难事。那么，带有明星脸的充气娃娃将会是宅男们的最爱，也可以把痛恨的人做成 3D 沙袋。但问题也就来了，用户的行为是否违法呢？2D 照片毕竟只是一张纸而已，而一模一样的 3D 模型着实有些吓人。

如今，3D 打印已经成为一种潮流，并开始广泛应用在设计领域，尤其是工业设计、数码产品开模等，可以在数小时内完成一个模具的打印。在国内，已经有很多企业进入这一市场。如此轻松就能克隆任何东西，这也引起业界担忧：用这种打印机克隆作品，就像随意在网上盗版一个电影一样简单。这种打印机一旦普及，无疑会给版权保护带来很大挑战。

3D 打印技术轻松、便捷的复制功能不仅给侵权盗版者提供机会，更为严重的是将加快侵权效率、降低侵权门槛、扩大侵权范围，这些问题正在引起国际社会的共同关注。据了解，著名的《数字千年著作权法案》对 3D 打印物品的相关问题十分关注。

然而，如何判断 3D 打印是否侵权，并不是一件简单的事，目前相关法规并不完善。目前 3D 打印主要通过以下 3 种方式进行：1）从 3D 到 3D，即将 3D 数字化模型打印成 3D 物品；2）从文字到 3D，即根据一段文字描述，如球体，半径为 5cm，颜色为蓝色等，进而打印出对应的 3D 物品；3）从 2D 图案到 3D，即将 2D 平面图案进行 3D 重建并打印成 3D 物品。现在的问题是，这 3 种打印方式是否属于著作权法保护的“复制”呢？

首先，从 2D 图纸到 2D 图案或是从 3D 模型到 3D 物品，都属于典型的著作权法意义上的复制，哪怕是缩印、扩印等改变比例的方式，都不影响复制的成立。因此，这种未经作者许可方式所进行的复制将可能构成侵权。

其次，从文字到 3D 的方法，一般不会认定为著作权法上的复制。著作权法保护的是“表达”，而文字与 3D 物品属于两种不同形式的表达方式，所以不涉及彼此复制的问题。因此，这种 3D 打印方式一般也不涉及侵权问题。

需要讨论的是，从 2D 平面图案到 3D 是否属于复制？我国著作权法对此问题避而未谈，实践中争议颇大。比如前面提到的将某位心仪女生的 2D 照片私自打印成 3D 实体模型，是否属于复制？因此在 3D 打印时代，有必要在立法中明确“复制”的具体方式，以便保护著作权人的合

法权益。

在讨论 3D 打印与知识产权侵权问题时，还必须关注到“合理使用”的问题。对于那种仅仅为了个人使用而少量复制的行为，会被认定为合理使用，从而被排除在侵权范围外。如果是这样的话，我们可以设想未来社会，很少有人会花费高额价格去购买“知名商品”，恰恰相反，人们更愿意花费低廉的成本购买原材料，在家里打印所需产品。众多消费者如此“合理使用”的结果，对于商家绝不是利好消息，这使“合理使用”将从根本上妨碍或者动摇经营者利益。

当然，对于版权和专利保护，要掌握一个合适的度，避免“一抓就死，一松就乱”。实际上，个人 3D 打印机之所以现在这么如火如荼，很大程度上归功于 FDM 工艺的专利到期，这才使得广大 3D 打印机厂商近几年如雨后春笋般涌现出来。FDM 工艺一般采用塑料为材料，精度、强度、耐用性离真正的产品还有距离，引起了很多用户的抱怨。而令人兴奋的是，2014 年 1 月，可打印金属材料的激光烧结工艺的专利保护期届满，这也被视为推动 3D 打印事业腾飞的利好消息。



提示：专利是有保护年限的，如发明专利保护期为 20 年，实用新型和外观设计保护期为 10 年。超过了保护年限，专利将进入公有领域，人人皆可免费使用。

音乐行业的版权保护案例可以给 3D 打印提供一个很好的参考。被美国唱片业协会（The Recording Industry Association of America）起诉过的用户多达 3.5 万人次，理由是非法在线共享音乐。但这种收效甚微，最后不得不在 2008 年做出改变，只对重大侵权事件提起诉讼。

9.9.3 社会伦理的思考及技术层面解决

前面我们已经谈到 3D 打印可能被犯罪分子用来制造枪支或是伪造货币，也可能引发山寨盗版泛滥、侵犯知识产权。除了这些公共安全和版权问题，实际上 3D 打印技术的发展对目前的社会伦理也有挑战，比如我们打印的材料不用塑料，而是活生生的细胞，达到了生物层面，人们可以克隆出一些器官甚至人，这会引发很大的社会伦理问题。

提取人体的干细胞制成“生物墨水”，用它们来将器官打印出来，这在生物学家们看来，并不是遥不可及的，实际上目前已有一些显著的进展。如果未来真正实现，器官捐献将不再需要，人类将会摆脱疾病、残疾。但是，如果有人想打印一双翅膀安在自己身上呢？还有人把一头猪的脑袋活生生地挂在自己腰间呢？或者，甚至有人把某个明星的脸打印在自己的后脑勺呢？这些邪恶的想象如果成真，究竟会出现超人还是怪物？你能接受你的家庭成员尝试这些新事物吗？所有这些新的伦理问题，究竟应该如何设定规矩和划清界限？

此外，3D 打印让设计不再是设计师的专利，让生产不再是工厂的专利。人人都可以生产东西。随之而来的就是设计海洋、创意疲劳和产品的狂轰滥造。环境保护者要头痛了，在自由想象与创造欲望的驱动下，随心所欲地涂鸦将带来灾难性的后果，各种耗材被民间大量使用。大量打印品的出现并快速地更新换代，会对环境造成不小的污染。尽管在制造业领域，3D 打印的出现会大大节省成本、节约资源。但普通百姓的大量使用，对资源的消耗将会有过之而无不及。

那么，我们能不能在技术层面避免以上种种可能的副作用呢？我们可以从技术层面上对 3D 打印这种几乎万能的制造能力进行一个限定和约束。比如就像音乐和电影业一样，我们可为制造业也设计一套**数字版权管理（DRM, Digital Rights Management）**系统。前微软 CTO 纳森·梅

尔沃德经营的知识产权风险投资公司，已被授予管理 3D 打印领域的“物品生产权”。该技术能在多大程度上防止未经授权的物品被复制，以及是否会影响到用户的打印体验，让我们拭目以待。比如 3D 打印机在打印之前自动分析形状，跟数据库中的版权保护文件进行比对，发现大部分雷同则拒绝打印。当然，对于枪支也是类似的，一旦发现 3D 图纸极有可能是枪支，则监管软件让打印机拒绝打印。在生物打印方面，类似地，会有更加严格的监管程序和技术手段，来阻止克隆其他人的器官或者随意打印未经许可的生物体。

总之，3D 打印机遇与风险并存。美国公共知识宣传组织的专职律师迈克尔·温伯格，在其论文《如果不搞砸的话，这将会很棒：3D 打印、知识产权以及下一个伟大的突破性技术之战》写道：“任何进步和那些受其鼓舞的人，不应该为了那些害怕改变的人而停止前进的脚步”。

9.10 3D打印3D打印机自己：遗传与升级

一个细胞通过不断的自我复制、分裂，可以产生无限多的细胞，正所谓“子子孙孙无穷匮也”。3D 打印机实际上可视为具有自我复制能力的机器人，可以通过打印自身的零部件，再将这些零部件组装出一台新的、一模一样的 3D 打印机。

我们在第 3 章 3.1 节“RepRap：开源 3D 打印机的鼻祖和奠基石”中已提到，第一台个人 3D 打印机 RepRap 的含义就是“复制和进化”，目前发布的几个版本的名称，如“达尔文”、“孟德尔”、“赫胥黎”也正取的是著名生物学家们的名字。我们在那一节还给出了 RepRap 打印机复制出一台一模一样的“儿子” RepRap 打印机的照片，以示所言非虚。

下面是个概念图，曾被作为愚人节的一个玩笑，以展示 3D 打印机可以不断复制自身的缩小版，如图 9-30 所示。如果真的如此，一个社区只需购买一台 3D 打印机就足够了，让它不断地“生崽”去。

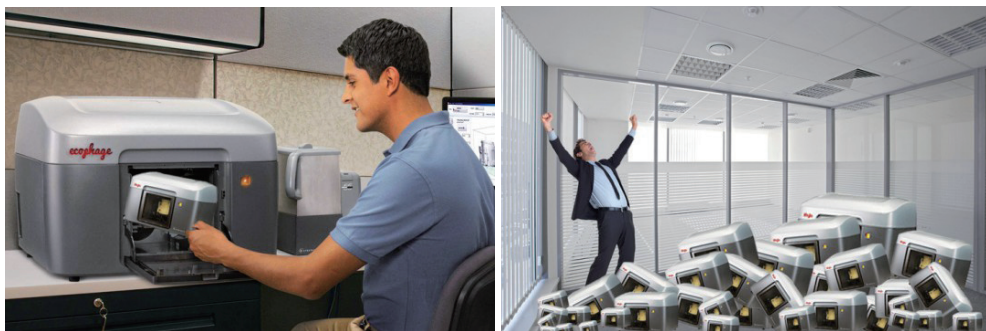


图 9-30 概念图，曾被作为愚人节的一个玩笑（图片来源：Doppelbock）

在科幻小说和电影中，可自我复制机器人的出现往往会在一定程度上制造恐慌，比如艾萨克·阿西莫夫的科幻小说和好莱坞“终结者”系列影片中的机器人，均拥有自我复制能力，它们为人类带来了可怕的灾难，但在目前看来，机器人距离那样的“逆天”能力其实还很遥远，而且也未必会对人类做出“狼顾”的行为，如图 9-31 所示。

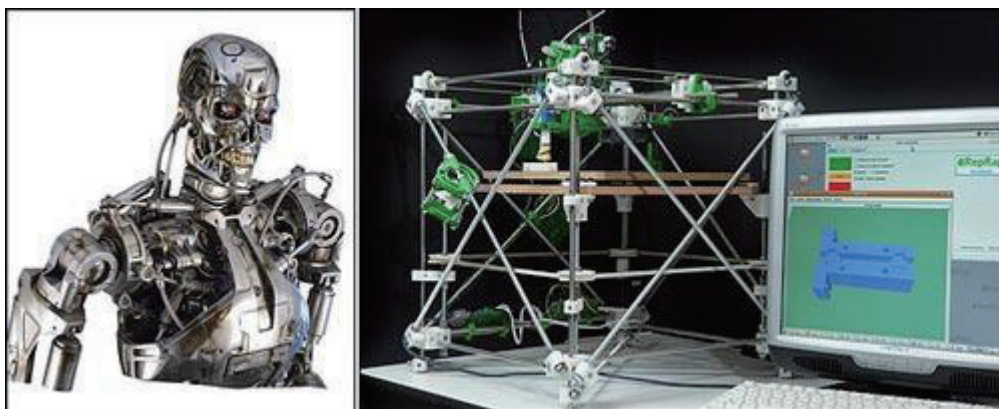


图 9-31 右图为可自我复制的机器人 RepRap，但要能达到左图终结者那样还很遥远
(图片来源: RepRap)

研制一种能够制造神经网络和大脑的机器人是 RepRap 小组的下一步工作，小组负责人 Adrian Bowyer 说：“下一代机器人将有能力制造电路，我们已通过实验证明它的可行性。我的一名学生添加的打印头便可以制造半导体”。

康奈尔大学霍德·利普森 (Hod Lipson) 博士领导的研究小组则更进一步，除了制造一台可利用“智能砖块”进行自我复制的机器人外，还制造了一个可以进行自我设计、安装甚至自杀的机器人。虽然尚无法制造出“终结者”系列中的液态金属机器人，但从某种程度上说，塑料机器人称得上液态金属机器人的前身。此外，利普森还研制出可进行自我修复的机器人——如果发现自己少了一条腿，就会发明一种新的走路方式。

世界上其他研究人员则致力于赋予机器人“自我维持”的能力，让它们通过摄入食物获取能量，如胡萝卜和有机肥料。英国布里斯托尔机器人实验室研制出了一款名为 Ecobot II 的机器人，它可以消化死苍蝇或者苹果等植物物质。梅尔赫什说：“这款机器人使用的是空气中的氧而不是外部化学制品，你也可以这样来理解，它是一个会呼吸的机器人。”目前该研究小组正研制一种可以产生“大便”的机器人，即有能力排除产生的废物。小组负责人坦言：“对于机器人，我们仍有很多事情可以去做。”

实际上，3D 打印机如果真的要变成智能机器人，可以遗传甚至自我升级，最关键的还是要产生类似于人类的智能。具体请参考第 4 章“3D 智能数字化：3D 打印的孪生兄弟”中的第 4.6.3 节“深度学习：像人脑一样深层次地思考”。

9.11 3D打印的经济模式：利基与长尾效应

“长尾理论”始作于美国《连线》杂志前主编克里斯·安德森。如图 9-32 所示就是著名的长尾示意图，曲线横坐标是产品受欢迎（越往左越热门，越往右越冷门）的程度，纵坐标则是相应的销售数量。从图中可以看到，最受欢迎的少部分产品，即图中左侧的“Body”（主体），品种不多但销量很大；右侧就是“Long Tail”（长尾），每个产品销量不多，但是因为产品种类多（也

即尾巴很长)，总的销量以及利润却可与头部相媲美，这从图中蓝色部分的总面积与红色部分的面积基本相当可以看出。

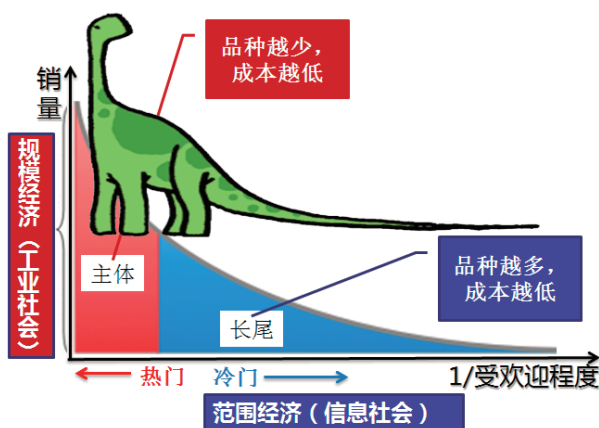


图 9-32 长尾 (Long Tail) 示意图

以互联网上的歌曲下载为例，几首大热门歌曲被下载了无数次，而随着曲目流行度的降低下载量陡然下跌。但有趣的是，它一直没有坠至零点。在统计学中，这种形状的曲线被称作“长尾分布”，因为相对头部来讲，它的尾巴特别长。这便是“长尾理论”的来历。

为了发挥人力资本的规模效应，大公司一般会专注于市场规模较大的“主体”，这部分产品种类不多，但被主流人群所需求，因此可以大规模生产，将成本降得很低。而众多满足个性化需求的“冷门商品”则被大公司忽视，这些“长尾”产品市场细分程度较高，种类繁多，而且利润丰厚，潜在市场规模的总量足以与主流市场匹敌。我们将这些高度细分的小市场称为**利基市场 (Niche Market)**，其共同组成了长尾市场。随着 3D 打印技术的出现和不断发展，个性化设计到制造的过程被大幅度简化，以个体化创意到网络平台化设计制造的商业模式将会得到有力的促进。这些产品以满足利基市场需求为目标，旨在攫取长尾部分的丰厚利润。



提示：“利基”一词是英文“Niche”的音译，意译为“壁龛”，有拾遗补缺或见缝插针的意思，这里指在大公司的布局空隙处见缝插针。长尾市场有时也统称为“利基市场”。

Shapeways 允许用户在其网站上设立商店，出售产品。这些用户将自己各种千奇百怪的创意成功转化成各种千奇百怪的商品，满足了利基市场的需求。按克里斯·安德森的理论，这些个体可以统称为“创客”(Makers)，而他们将获取商品长尾价值、发展利基市场的驱动因素。同理，在软件领域既有微软、IBM 这样的巨无霸，也有 App Store 上为数众多小 App 的开发者，在面向特定人群的利基市场上获取着丰厚的利润。

亚马逊网站被视为长尾理论的成功实践：在书、影音等消费领域，依靠“无尽的网上货架”，所有产品都有机会进行销售，消费者无尽的选择需求都能得到满足，这些长尾商品“积少成多、积沙成塔”之后往往可以累积起一个足够大的量，与主流热门商品相匹敌。网络可以做到的是哪怕这个网站一年只卖出一本极其罕见的书给消费者，它的营销成本并不显著增加，甚至根本不增加。但是实体店却做不到，因为时间、成本和空间都不允许他们这么做。

长尾得以存在有如下几个前提:经济发展到富足阶段,热销商品向小众/细分/利基市场转变,用户找到冷门产品变得容易,众多小市场聚成一个大市场。

- 长尾理论讲述的是**富饶经济学**。而传统的**稀缺经济学**假设这个世界是一个产品、资源、物质极度匮乏的世界,因此只有生产那些最热门的商品来满足大众的日常需要,并通过庞大的销售数目来降低售价、提高利润。富饶经济学则假定这个世界是一个产品种类、物质极大丰富的世界,提供的产品远比以前要多。
- 随着经济的发展,人们生活水平的提高,用户不再仅仅需要那些最“热门”的大路货,相反,他们也十分关心那些有个性的**小众商品**。消费者越来越青睐**多样化市场**。
- 由于搜索引擎和推荐系统(请参考第4章4.6.2节“大数据背景下的个性化推荐系统”)的功劳,找到这些冷门的小众产品远比从前要容易,这样销售额在大热门和小市场之间的分配更加均匀了。
- 尽管每一种小众商品的销售量和利润不足以支持一个店铺或一家企业的运营,但是当小众商品的数量达到一定程度后,加上单件的利润都很高,则店铺或企业的总利润将十分可观。

传统的市场曲线是符合**80/20定律**(又名**二八定律**、帕累托定律、最省力的法则、不平衡原则)的,为了抢夺那带来80%利润的畅销品市场,我们厮杀得天昏地暗。今天,尽管我们仍然对大热门着迷,但它们的经济力量已经今非昔比。消费者们已经散向了四面八方,市场已经细分成了无数不同的领域。因此,就算20%的产品能带来80%的销量,我们也没理由不去经营其他那80%的产品,因为存货成本极低。从某种意义上讲,长尾理论是**蓝海战略**的延伸(因此图9-29中的长尾部分用蓝色标注,以形象揭示利基市场不存在血腥竞争,尽可在蓝色大海中自由游弋),因为蓝海战略讲究的是要避开红海的血腥竞争来使得企业在市场竞争中保持优势。

实际上,长尾效应揭示了传统工业社会的规模经济向当代信息社会的范围经济的转变。在**规模经济(Economies of Scale)**社会中,品种越少,越便于大规模生产,成本也就越低。而在**范围经济(Economies of Scope)**中,品种越多,分摊下来单件成本越低。以电信通信业为例,每年在通信设备和线路上的成本投入都是固定的,但闲着也是闲着,可通过开发各种服务品种,如将电话、电报、传真、电视、宽带、3G,以及各种名目繁多的套餐服务,极大地分摊每种业务的成本。再以3D打印为例,卖茶杯的店铺可同时提供100多种形状和颜色,由于3D打印任意成型,无须制作专门的模具,所以总的研发成本并没怎么增加,分摊下来就大大降低了每一个品种的单位成本。

在现在的网络经济中,亚马逊利用长尾特征和“**即需即印**”模式来解决书籍的存货虚拟化问题,已经是最接近现实世界的手段了。但如果加上了3D打印呢?理论上,网络平台(如亚马逊、淘宝、阿里巴巴、京东等)几乎可以卖任何东西,当我们把3D打印看作网络经济向现实世界全面渗透(即从“比特世界”到“原子世界”的过程)的主要通道时,我们就可以体会到奥巴马大力发展3D打印的兴奋了——毕竟美国是当前世界上网络经济最发达的国家,而3D打印天生就具备“即需即印”的属性。

在图9-29中,“主体”意味着单一性的大规模生产,而“长尾”意味着差异化、多样性的小批量生产。今天的市场上二者并存,但后者代表着未来。安德森认为定制和小批量才是制造业的

未来。中国要靠中小企业的长尾战略缔造国家竞争优势，在理论上是可能的。但绝对的先决条件，必须是走技术创新的道路，否则就没有长尾最看中的东西——利润。

9.12 “中国智造”推动“全球第三次工业革命”

由“智能数字化制造”引发的第三次工业革命山雨欲来。2010年，中国大约有1.3亿人从事制造业，约占全球制造业工人的40%。同时，中国的竞争优势早已不限于低成本的普通劳动力，更有大量雇佣成本适中的中等技术人才、工程师、科研人员，以及可靠的供应链、完善的产业链条、高度适应性的巨大产能和本身巨大的市场。

3D打印之所以配得上“工业革命”这一字眼，是因为将从两方面对制造业产生深远的影响。

- 首先，3D打印影响了传统制造企业的生产方式。基于这一技术的应用，微小企业应对市场多元化需求和不确定性的柔性生产能力将会进一步提高。在经历了以福特为代表的标准化制造，和以丰田为代表的精益生产之后，制造企业在未来可能会步入自由生产的阶段。制造业整体则有望经历一次从“规模生产”到“规模定制”的转变。
- 另一方面，3D打印将创造以“智能云网”、“个人智造”、“家庭智造”、“网络社区智造”为代表的新工业模式。由于3D打印对于制造流程的简化和数字化程度的提升，设计和制造的技术门槛被大幅降低，从而使非专业人员同样可以实现产品的设计和制造，激发更多的个性化设计。同时，专业化的智能技术将以云端智能服务和“创件”的形式封装给缺乏相关领域知识的产品开发者，使得后者可以在很短的时间迅速打造跨学科领域的专业智能化产品，此外，借助网络社区平台的信息共享、协作创新、在线交易、口碑营销功能，将个人创意转化为可以创造赢利的实际产品并在线销售，打开了专业制造向个人智造转化的大门。

当前，中国正处于从“中国制造”向“中国智造”迈进的重要时期，3D智能数字化及3D打印技术可以让国内的设计师和工程师从产品制造工艺的束缚中解放出来，更加专注于产品本身的智力创造，大跨步进入想法到产品（Mind to Product）的“所想即所得”全新智造时代。同时，智造时代还具有两重属性：既是制造业，也是服务业，而我国在这两方面都具有巨大的人力成本优势。3D智能数字化和3D打印的产业化无疑将为促进我国传统产业升级、彻底摆脱长期处于制造业产业链底端的尴尬局面发挥十分重要的推进作用。

中国要成功地将“中国制造”转型升级为“中国智造”，并以此推动“全球第三次工业革命”，除了牢牢把握以智能数字化为核心的信息技术，同时还需要全面推进新能源技术、新材料、先进制造（包含3D打印）、生物技术、航空航天技术等方向的发展，这些都是新工业革命的重要组成部分。

9.12.1 新工业革命之“永不枯竭的绿色能源”

中国作为世界头号制造业大国，严重依赖能源。19世纪，英国引领了第一次工业革命，推进了煤炭的使用并创造了日不落帝国。20世纪，美国引领了第二次工业革命，推进了石油的使用。21世纪，中国正处于一个煤炭、石油、天然气和核能时代，这些能源不但费用越来越高，而且

对环境造成了污染，对经济发展的推动作用越来越不可持续。第三次工业革命的基础将是可再生的绿色能源。

以可再生能源的储量而言，中国就是新能源领域的“沙特阿拉伯”，拥有世界上最丰富的风力资源，也是世界上太阳能最为丰富的国家之一，生物能与地热能也相当丰富。其中，利用太阳能的最佳方式是光伏转换，就是利用光伏效应，使太阳光射到硅材料上产生电流直接发电。以硅材料的应用开发形成的产业链条称为“光伏产业”，包括高纯多晶硅原材料生产、太阳能电池生产、太阳能电池组件生产、相关生产设备的制造等。

在政府的支持下，我国光伏产业发展迅速，已形成较为完整的光伏制造产业体系。2009年中国内地多晶硅产量超过了2万吨，太阳能电池产量超过了4000兆瓦，连续3年成为全球太阳能电池的第一生产大国。然而，95%的产品出口，说明中国并不是在为自己生产。中国目前使用的能源中只有0.2%是可再生能源，我们不惜承受着第二次工业革命时代“高污染”（多晶硅电池制造属于高耗能和高污染产业）的昂贵代价把第三次工业革命的基础设施传至世界上的其他国家，使得他们能进入下一阶段。这不合理，但体现了一种“崇高的革命友情”！

可是，美国和欧盟似乎并不领情，2011—2012年先后对中国的光伏企业展开了“双反”调查，对来自中国的光伏产品征收31.14%~249.96%的高额反倾销税。同时，由于缺乏创新驱动，在核心技术领域没有话语权，我国的光伏企业只能挤在价值链低端进行恶性竞争，虽然仍可以靠量大取胜，但利润却很微薄。

中国制造业要真正做强做大，除了坚持自主创新之外，还必须加快全球化进程，在全球范围内参与市场竞争，整合优势资源。目前在欧债危机的影响下，欧美等发达国家经济发展陷入困境，一批拥有世界顶级科技的公司面临着重组的局面，而中国企业可以抓住机遇，通过并购全球优秀制造业企业，实现对国际领先技术的控制与中国落地应用。2012年，一家中国民营企业——鑫明光集团收购了美国太阳能上市公司Ascent Solar，后者是目前世界唯一的塑料基体的铜铟镓硒（CIGS）柔性薄膜太阳能电池量产供应商，其技术被美国《时代周刊》评为世界五十大大发明之一。这个例子说明中国已经在不断向“智造”迈进，并将不断深入到全球新工业革命的核心发展进程当中。

德意志银行2013年1月14日发布报告称，2014年中国太阳能发电装机容量将翻倍，中国将超越德国成为市场领头羊，中国也将由此跃升为全球第一大太阳能市场。无论是面对潜力巨大的内部市场，还是需求强劲的外部市场，中国正以自己的实际努力致力于“全球第三次工业革命”绿色能源的普及和应用。

9.12.2 新工业革命之“3D打印新材料”

3D打印已被公认将引发一场制造业革命。其实，3D打印并不是一项高深艰难的技术，它与普通打印的区别就在于打印材料。

可打印材料的稀少和昂贵是制约3D打印技术的瓶颈。3D打印需要的材料必须是高分子材料，加热或激光照射之后要具有流动性，能够变成流体、半流体，成型之后有能力立刻固化。拥有这样条件的材料目前并不多。

当前，以色列的 Objet（现已被美国 Stratasys 公司收购）是掌握最多打印材料的公司。它已经可以使用 14 种基本材料并在此基础上混搭出 107 种材料，并成功实现了多种材料的混合制造，这样免去了产品组装的步骤。但是，这些材料种类与人们生活的大千世界里的材料相比，还相差甚远。不仅如此，这些材料价格便宜的要每千克几百元，最贵的要每千克 4 万元左右。当然，根据作者对行业内部的了解，目前国外公司的销售价格属于暴利。而国内的材料价格就实在得多，如已经国产化的光敏树脂材料和热熔性塑料，品质上跟国际基本平齐，而价格仅为 1/5 左右。

3D 打印的发展也体现了材料的决定作用。当只能使用塑料、石膏的时候，3D 打印的应用主要局限在样品模型制作上；当它能使用钛合金的时候，就能直接打印出构件应用在飞机上。如果未来几年，土壤、岩石、细胞等都能成打印材料，再造一个大千世界也就成为了可能。因此，当中国生产的“物美价廉”的打印材料成批运出国门之时，也就是新工业革命席卷全球每个角落之日。

9.12.3 新工业革命之“先进制造及 3D 打印”

工业革命的基本定位在于制造，只有制造出实实在在的产品，人们的生活水平才能得到实实在在的改善和提高。而第三次工业革命与前两次工业革命不同，强调先进制造、高端制造、智能制造，即以数字化、智能化为核心，提升生产效率、技术水平和产品质量，并降低能源资源消耗。

中国政府正在加大对“高端装备制造业”的支持力度，已将其列入国家七大战略性新兴产业之一，而“智能制造装备”又是其重点发展方向。《国务院关于加快培育和发展战略性新兴产业的决定》中明确指出“强化基础配套能力，积极发展以数字化、柔性化及系统集成技术为核心的智能制造装备”，并出台了相关配套政策措施。按照《智能制造装备产业“十二五”发展规划》的要求，到 2015 年，我国智能装备制造业的销售收入将超过 1 万亿元，年均增长率超过 25%，工业增加值率达到 35%。本土化智能制造装备的国内市场占有率将超过 30%。力争到 2020 年，形成完整的智能制造装备产业体系。

3D 打印是先进制造、高端制造、智能制造的代表性发展方向。在高端制造领域，像 F22 猛禽战斗机和波音 787 大客机都需要钛合金结构件，如采用传统锻造和机械加工工艺，既耗时又费力，还需要数万吨级重型液压锻造装备进行加工，材料利用率通常不足 10%，也即浪费率高达 90%。但应用 3D 打印则彻底不同，与传统工艺相比，材料利用率和浪费率正好颠倒过来。

目前中国在 3D 打印技术领域已经与美国、欧洲等国际巨头“基本处于同一水平”。中国正在研发的 C919 大型客机，其机头工程样件就需要钛合金主风挡窗框，如图 9-33 所示。该部件从欧洲订购需要两年，每个部件的锻造仅模具费就要 50 万美元。而我国科研人员目前在大型金属整体构件直接制造等高端工业制造领域取得了“革命性”突破，如北京航空航天大学研究团队的激光直接制造技术，从制造零件到装上飞机仅用了 55 天，零件费用还不到其模具费的 1/5。

中国 3D 打印技术产业联盟以及中国 3D 打印研究院的成立，将集成国内研发力量（如北京航空航天大学、清华大学、华中科技大学、西安交通大学、西北工业大学等），重点开展医疗康复、航空制造、航天科技、汽车研发、生物制造等领域 3D 打印工艺、装备、材料、应用等的技术研发和产业转化。



图 9-33 C919 风挡在高速飞行时要承受巨大动压，窗框由钛合金 3D 打印制成（图片来源：新华网）

9.12.4 新工业革命之“3D 智能数字化创造”

以上几节，我们分别介绍了第三次工业革命的几个基础，如绿色能源、新材料、先进制造。在全章的最后一节，我们将讨论第三次工业革命的核心概念：数字化与智能化，这也是贯穿本书全部内容的一条主线。“数字化与智能化”在“3D 打印”上的着力点便是“3D 智能数字化创造”，赋予了“万能制造机”3D 打印以鲜活的灵魂、智能的大脑、创新的精神和自由的形式。

谈到创新，那么当前中国的创新能力在世界上处在什么位置？可能认为“与发达国家差距较大”的人会比持相反判断的人更多。然而世界著名的毕马威（KPMG）咨询公司给出的调查结果却可以说是出人意料。据英国《金融时报》2012 年 6 月 27 日的报道，由毕马威组织的一项面向计算机和电子等行业逾 650 名高管的调查显示，有 30% 的被调查者认为中国将在未来 4 年内成为最大的“全球创新热点”，排在第一位；美国得票率为 29%，排名第二；其后是印度、日本和韩国，得票率分别为 13%、8% 和 5%。实际上，创新分“基础科学原始创新”和“技术工艺改进创新”两种，在工业制造领域，真正对产品起决定性作用的还是后者，而中国在这方面的创新能力正突飞猛进。

通过本书第 2 章和第 3 章的介绍，相信大多数读者对于 3D 打印机的原理和结构已经有有了一个全面的了解。其实，对于 3D 打印机的使用者而言，既不需要去亲手设计一款新的 3D 打印机设备和控制系统，也无须自己研制 3D 打印机使用的材料。我们只是把 3D 打印机作为一个桌面制造工具来使用，希望打印出什么东西才是我们的关注点，也就是说，作为 3D 打印机的信息输入源，一个满足实际应用的“3D 数字化模型”才是体现我们智力创新的成果。而这，也正是本书所反复强调的重点所在。3D 打印的优势在于可以将这些任意复杂的个性化设计轻松制造出来，这无疑会激发我们每一个人的创造热情！

目前 3D 智能数字化技术在制造业，尤其是设计制造环节发挥着重要作用，已成为各国争夺行业制高点的竞争焦点。比如在医疗外科领域，如图 9-34 所示，骨科医生需要为患者设计最优化的植入体与接骨板，就只能依靠 3D 智能数字化技术。具有多年研究经验的上海交通大学教授王成焘提到，骨科手术治疗包括很多环节。首先要利用 CT 扫描数据建立 3D 数字化模型，呈现患者骨头内部的精确结构（图 A）。接着，根据个体的不同结构，利用智能软件设计个性化的

植入体（图 B）。所设计的植入体到底合适不合适呢？我们可以进一步通过智能算法来分析植入体对骨头各区域的受力影响（图 C 中不同的颜色代表着不同的受力大小），看力学性能分布是否合理。之后，我们就可以将模型 3D 打印出来，在真正的手术之前进行规划和反复演练，待熟练之后就可以正式实施，这样做的目的是可以让手术更加准确和精细。

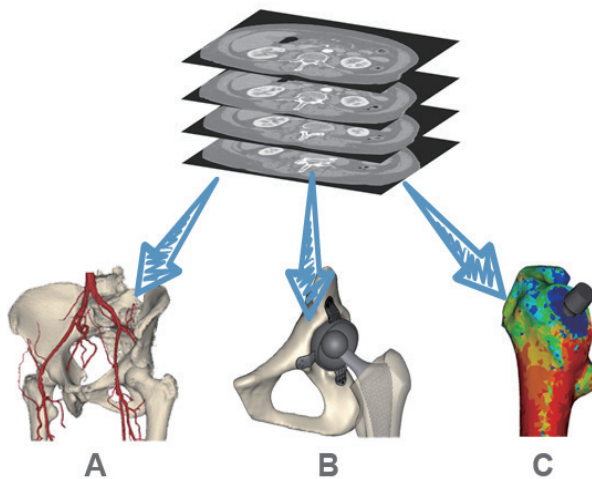


图 9-34 A：数字化 CT 扫描；B：智能化设计；C：分析受力性能（图片来源：Materialise）

3D 打印产业的迅猛发展必然会对 3D 智能数字化技术提出新的要求。过去，3D 打印通常面向模具制造、工业设计、医疗卫生等专业领域，都是聘请专业的技术人员做算法的开发与设计，并没有向大众普及的需要。但现在，3D 打印已经从高端工业走向大众化，这就需要软件设计也向大众化转型，这将会开启一个巨大的 3D 智能数字化产品消费市场。3D 智能数字化设计者可以把自己编写的专业软件以云端智能化服务的形式提供给普通用户和开发者，以降低 3D 智能数字化技术的应用门槛。同时，作为“智能云网”的重要组成部分，“云端智能服务”与“云制造”一样，也是一种可行的商业赢利模式。目前我国科研单位（如中国科学院、浙江大学、清华大学、北京航空航天大学等）在 3D 智能数字化方面已有很强的技术储备，已基本解决了相关的技术难点，只是没有资金实力形成功能完整的大型软件系统，因此可先针对需求较为单一的细分行业进行突破，并以本土化的优势服务于呈星火燎原之势的“个人智造”、“家庭智造”、“网络社区智造”。

可以看出，“3D 打印”不是一个单一和孤立的学科，实际上，仅凭单项技术是不可能、也无法掀起一次工业革命的。3D 打印已逐渐发展成为一个新兴的交叉学科和丰富的泛技术群，包括信息技术、先进制造技术、新材料技术、新能源技术、生物技术、航空航天技术和海洋技术等。因此，2014 年，“中国 3D 科技创新产业联盟”在北京中关村筹备成立，旨在加强 3D 智能、3D 打印、3D 数字化、3D 识别、3D 音视频、3D 建筑等各个领域之间的交流与合作，联盟成员包括中国科学院、中国工程院、清华大学、北京大学、同济大学等几十所著名院校及国内外企业、投资集团。这个以 3D 专业为基础的非营利组织定位于中国 3D 产业发展国情，通过技术路径、商业模式、服务标准的创新，促进中国 3D 科技产业的健康良性发展。

可以看出，中国正在加大对这些关键技术领域的投入和支持力度，以期通过占据核心产业制高点，形成“中国智造”新模式，推动全球第三次工业革命向有利于中国的方向发展。

第10章

道：数字智能的最优化及相关数学方法

《庄子·天下》有云：“不离于宗，谓之天人”。《荀子·儒效》又云：“千举万变，其道一也”。清代的谭献则在《复堂类稿·明诗》中更直接地点出：“万变而不离其宗”！确实，3D 打印和 3D 智能数字化发展日新月异，大量的新方法层出不穷，是一本书的有限篇幅远远不能穷尽的。那么，它们的“宗”或者“道”到底是什么？本书的回答是：数学原理（和对应的几何意义）。因此，你无须奇怪最伟大的物理学家牛顿，将他那本划时代的物理学著作命名为《自然哲学的数学原理》。同样，你也可以明白为何本书将数学作为最后一章，作为“压箱底”的镇关之宝。

实际上，在数字化、智能化创造时代，数学是必不可少的。例如，3D 打印的“先切片再累积”的工作原理，本质上用的就是数学中的微积分思想。在 3D 智能数字化中，数学用得就更多了。而且非常有意思的是，各种方法给出的数学形式往往最终都归结于最优化问题的求解：即给定一个目标函数，使得这个函数能够取得最小值。例如：

- 在 3D 打印涉及的材料科学领域，最小化能量需要用到最优化方法。
- 在给定载荷和环境条件下，我们可用外点惩罚函数法来优化设计 3D 打印机的某些机械部分，以提高可靠性、减小体积、降低成本。
- 对于第 6 章 6.2.3 节中的形状编辑，我们要尽可能刚性地保持形状细节，以最小化扭曲误差，这又涉及最小二乘问题的求解。
- 在 6.4.2 节中，支持向量机的数学模型可以归结为一个二次规划问题。
- 在 6.8 节中，我们要设计一个最优的内部结构，既要尽可能地使质量和耗材最省（高优先级），又要使得结构尽可能地简单和不冗余（低优先级），这涉及多目标规划问题。

下面对本书所涉及的常用数学方法做一个介绍，包括最优化方法和贝叶斯概率方法。同样，笔者仍将以最通俗易懂的话，把这些原本深奥晦涩的数学讲述清楚，你只需有最基础的高等数学知识即可。当然，如果你想进一步了解更细节的证明和推导，可参考参考文献^{[17][65~67]}。本章是 3D 打印科研人员和 3D 智能数字化算法开发人员必备的高阶理论基础。

10.1 最优化理论的基本常识

关于最优化，著名的大数学家 Euler（欧拉）曾说过：“Für, da das Gewebe des Universums am vollkommensten und die Arbeit eines klügsten ist Schöpfers, nichts an findet im Universum statt, in dem irgendeine Richtlinie des Maximums oder des Minimums nicht erscheint.（由于宇宙组成是最完美也是最聪明造物主之产物，宇宙间万物都遵循某种最大或最小准则）”。这实际上就是说最优化无处不在。实际上，根据达尔文的进化论，大自然的万物遵循着“优胜劣汰”的法则，即在给定约束条件下（如气候、能源、地理条件），朝着最适应的方向进化。比如，猎豹所进化出的身体结构使它奔跑起来具有最优的爆发力；你也同样不必惊奇海豚的外形是光滑优美的曲面，而不是任意生成的坑坑洼洼的噪声曲面。

最优化（Optimization）理论，或称为**数学规划（Programming）**、**运筹学（Operations Research）**，指研究数学上定义的问题的最优解，一般可归结为对**目标函数（Objective Function）**（或称之为误差函数 Error Function、代价函数 Cost Function、损失函数 Loss Function）求极值的问题：即对于目标函数 $f(x)$ ，找到一个极值点 x^* ，使得 $f(x^*)$ 最小（或 $-f(x^*)$ 最大）。

形象地说，最优化相当于盲人爬山，如图 10-1 所示。盲人爬山是为了登上山顶，而最优化是为了求取极小值或极大值。盲人在登山时，只知道脚底下的情况（如当前的所在位置、地面倾斜的坡度），而看不见其他任何地方的情况。最优化在求取极值时也跟盲人一样，只知道当前点的信息（如函数值大小、一阶导数梯度的大小和方向），但不知道其他点的信息。此外，除了一阶导数信息，还可查看二阶导数 Hesse 矩阵，若矩阵正定（负定），则当坡度为 0 时，地面的弯曲是下凸的（下凹），即是谷点（峰点）。

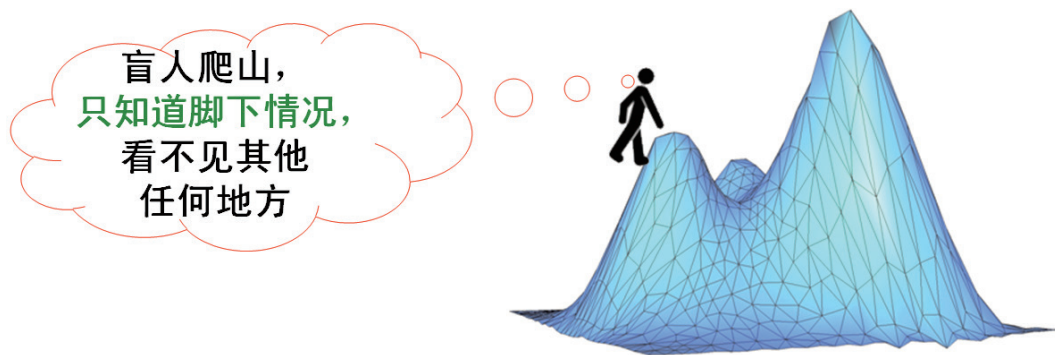


图 10-1 最优化相当于盲人爬山

从图 10-1 中还可以看出，如果有多个山峰，盲人登上的山峰未必是最高的那座山峰。计算机也一样，对于多峰这种非凸函数，有可能求取的只是局部极值，而不是全局极值（最值）。

10.1.1 从凸集和凸函数开始说起

为了保证求取的极值不仅是局部极值，同时还要是全局极值，我们引入凸集、凸函数和凸优化的概念。

凸集

设集合 $\mathbf{X} \subset \mathbf{R}^n$ ，若对 $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}$ ，及任意一个数 $\lambda \in [0, 1]$ ，都有 $\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \in \mathbf{X}$ 成立，则称 \mathbf{X} 为凸集。其几何意义表示为：如果集合 \mathbf{X} 中任意两个元素连线上的点也在集合 \mathbf{X} 中，则 \mathbf{X} 为凸集。如图 10-2 所示，左图为凸集，右图不是。

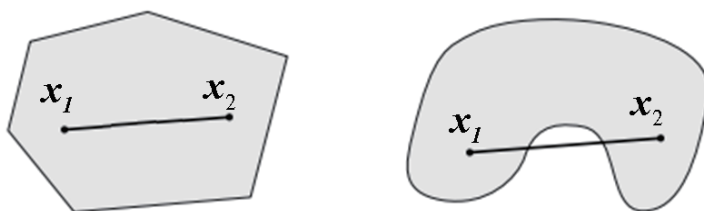


图 10-2 凸集与非凸集

凸组合

有 K 个点 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$ ，若存在非负的实数 $\lambda_1, \lambda_2, \dots, \lambda_K$ ，即 $0 \leq \lambda_k \leq 1$ ($k = 1, 2, \dots, K$)，且总和为 1，即 $\sum_{k=1}^K \lambda_k = 1$ ，则线性组合 $\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \dots + \lambda_K \mathbf{x}_K = \sum_{k=1}^K \lambda_k \mathbf{x}_k$ 称为点 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$ 的凸组合 (Convex Combination)。显然凸集 \mathbf{X} 中任两点 $\mathbf{x}_1, \mathbf{x}_2$ 连线上的任意点都是 $\mathbf{x}_1, \mathbf{x}_2$ 的凸组合。

如果不限定 λ_k 为非负，但总和仍为 1，则称为仿射组合 (Affine Combination)。如果不限定 λ_k 为非负，也不限定总和为 1，则称为线性组合 (Linear Combination)。

凸函数

设 $f(\mathbf{x})$ 为定义在凸集 \mathbf{X} 上的函数，若对任何实数 $\lambda \in (0, 1)$ 以及 \mathbf{X} 中任意两点 $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}$ ($\mathbf{x}_1 \neq \mathbf{x}_2$)，恒有：

$$f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2)$$

则称 $f(\mathbf{x})$ 为定义在凸集 \mathbf{X} 上的凸函数。以一维的情况为例，其几何意义为曲线上任意两点的连线总在曲线的上方，如图 10-3 左边所示。可以看出，(下)凸函数形如 \cup ；反之，(下)凹 (上凸) 函数形如 \cap 。若 $f(\mathbf{x})$ 为凸函数，则显然 $-f(\mathbf{x})$ 必为凹函数。仿射函数 $y = ax + b$ (线性函数 ax 加一个常量 b) 既是凸函数，又是凹函数。

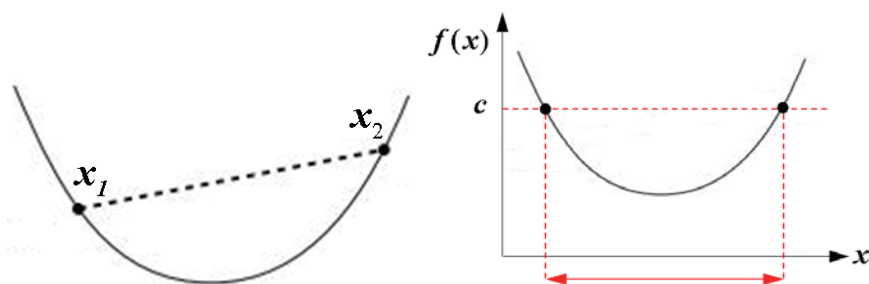


图 10-3 左：凸函数；右：水平集

当然，靠凸函数的定义来判断目标函数是不是凸函数，在实际中就比较难了，通常都是用目标函数的梯度 $\mathbf{g}(\mathbf{x})$ （一阶导数）或二阶导数 $\mathbf{H}(\mathbf{x})$ （Hesse 矩阵）来判断，人们已经得到了很多定理，例如，**一阶判别条件**为：任意一点处的切线增量 $\mathbf{g}(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1)$ 不超过函数的增量 $f(\mathbf{x}_2) - f(\mathbf{x}_1)$ ；**二阶判别条件**为：在定义域内任意一点 \mathbf{x} 上目标函数的 Hesse 矩阵 $\mathbf{H}(\mathbf{x})$ 半正定，则其为凸函数。

水平集

水平集 (Level Set) 是指具有某个相同函数值 c 的变量 \mathbf{x} 的集合：

$$\{\mathbf{x} \mid f(\mathbf{x}) = c\}$$

其中 c 是常数。当 \mathbf{x} 为二维时，又被称为水平曲线、**等高线 (Isocontour)**、**隐式曲线 (Implicit Curve)**。

还有一个定义叫作**子水平集**：

$$\{\mathbf{x} \mid f(\mathbf{x}) \leq c\}$$

以一维的情况为例，指的是如图 10-3 右边所示的两个水平点之间的所有点的集合。**在最优化理论中**，“水平集”一般指的就是“子水平集”。

凸优化 (凸规划)

若 $f(\mathbf{x})$ 为定义在凸集 \mathbf{X} 上的凸函数，则 $f(\mathbf{x})$ 的局部极小点就是它在 \mathbf{X} 上的全局极小点。由此我们引入**凸优化 (凸规划)** 的定义：

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = 0 \\ g(\mathbf{x}) \geq 0 \end{aligned}$$

其中，*s.t.* 是 subject to 的缩写，表示“以……为约束”、“以……为条件”、“假定”、“满足于”的意思；目标函数 $f(\mathbf{x})$ 是凸函数；**可行域**（满足所有约束方程的点 \mathbf{x} 的集合）是凸集，即不等式约束 $g(\mathbf{x}) \geq 0$ 是凹函数（或 $g(\mathbf{x}) \leq 0$ 是凸函数）、等式约束 $h(\mathbf{x})$ 是线性（仿射）函数。

驻点及鞍点

驻点 (平稳点)：一阶导数为 0 的点。它包括 3 种类型：极小值点、极大值点、鞍点。

鞍点：如图 10-4 左边所示为给出的马鞍形状，也即**马鞍凹下去那部分的中心点**，沿一个方向取得极大值，沿另一个方向却取得极小值。这个点虽然一阶导数为 0，但却“既不是极大值点也不是极小值点”的点称为鞍点 (Saddle Point)。

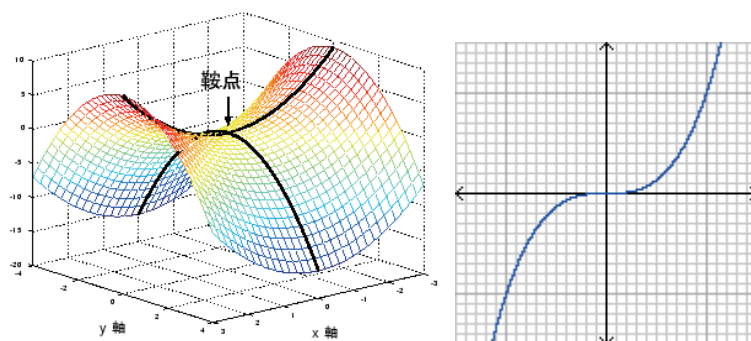


图 10-4 左：二元函数的鞍点；右：一元函数的鞍点（图片来源：tau.doshisha.ac.jp/~kon/、维基百科）

10.1.2 无约束优化与约束优化

无约束优化

对于 n 维变量 $\mathbf{x} \in \mathbf{R}^n$ ，无约束优化是指求解：

$$\min f(\mathbf{x}) \quad (1)$$

即找到一个极值点 \mathbf{x}^* ，使得目标函数值 $f(\mathbf{x}^*)$ 最小。

约束优化

约束优化指在 (1) 的基础上加入了一些约束条件：

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = 0 \\ g(\mathbf{x}) \geq 0 \end{aligned} \quad (2)$$

其中满足所有约束方程 (2) 的点 \mathbf{x} 所组成的集合，称为可行域 (Feasible Region)。

10.1.3 线性规划与非线性规划及其对偶 (Dual) 形式

线性规划与非线性规划

约束优化的情况下，当目标函数 (1) 和约束函数 (2) 均为线性函数（如： $h(\mathbf{x}) = ax_1 + bx_2 + c$ ）时，问题就称为线性规划 (LP, Linear Programming)。而当目标函数和约束函数中至少有一个是变量 \mathbf{x} 的非线性函数（如 $f(\mathbf{x}) = ax_1^2 + bx_2$ ）时，问题就称为非线性规划 (Non-Linear Programming)。

线性规划目前已经有成熟的通用求解方法，如单纯形法 (Simplex Method)、内点法等。在几何上，线性规划可理解为求凸多面体最低的点。Dantzig 提出的单纯形法本质上就是每次从凸多面体的一个顶点走到相邻的一个更低的顶点而逐步找到最低点的方法，方法简单、直观，但复杂度是指数级的。Karmarkar 提出的内点法的基本思想是从凸多面体的内部而不像单纯形法那样在边界上去逐步靠近最优解，此方法可降低到多项式复杂度。

此外，还有几类特殊的线性规划。比如要求最优解必须为整数（如所需机器的台数），则称为**整数线性规划（ILP）**，解法有**分支定界法（Branch and Bound）**、**割平面法**等。更特殊地，如果最优解必须为 0 或 1（是或否），则称为**0-1 整数线性规划问题**，典型的有**背包问题**、**集合覆盖和布点问题**等，解法有**隐枚举法**等。

比 0-1 整数线性规划更特殊的是“**指派问题**”，指派 n 个人分别去完成 n 个任务（每人只指派 1 项任务，即对其他任务的指派标签都是 0），由于个人能力的多样性，每个人对每个任务所需的开销各不相同，求最佳指派以使总开销最小，解法有**匈牙利算法（Hungarian Algorithm）**。

此外还有几种其他的特殊线性规划问题，如**运输问题**、**投资问题**、**配料问题**等，有一些专门的高效解法，如**表上作业法**。

线性 / 非线性规划的对偶形式

首先给出**线性规划的对偶（Dual）形式**：

$$\begin{array}{ll} \max & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \min & \mathbf{y}^T \mathbf{b} \\ \text{s.t.} & \mathbf{y}^T \mathbf{A} \geq \mathbf{c} \\ & \mathbf{y}^T \geq \mathbf{0} \end{array}$$

可以看出：左边**原问题**求极大值，右边**对偶问题**则求极小值，且两个问题最优解的目标函数值必相等。**对偶问题的对偶是原问题**。最大化问题的任意一个可行解的目标函数值都是其对偶最小化问题的目标函数的下界；最小化问题的任意一个可行解的目标函数值都是其对偶最大化问题的目标函数的上界。原问题和对偶问题的最优解满足**互补松弛性关系**。在经济学领域，对偶解就是影子价格，大小反映了资源在系统内的稀缺程度。

接着给出**非线性规划中二次规划的对偶（Dual）形式**：

$$\begin{array}{ll} \min & \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{g}^T \mathbf{x} \\ \text{s.t.} & \mathbf{Ax} \geq \mathbf{b} \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \max & (\mathbf{b} + \mathbf{A}^T \mathbf{H}^{-1} \mathbf{g})^T \boldsymbol{\lambda} + \frac{1}{2} \boldsymbol{\lambda}^T (\mathbf{A}^T \mathbf{H}^{-1} \mathbf{A}) \boldsymbol{\lambda} \\ \text{s.t.} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

其中 $\boldsymbol{\lambda}$ 为 **Lagrange 乘子**。我们可以把难解的原问题转到对偶问题，使其容易处理。

我们知道，**SVM（支撑向量机）也是求解二次规划问题**，其对偶形式为：

$$\begin{array}{ll} \min_{\mathbf{w}, \mathbf{b}} & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \max_{\boldsymbol{\lambda}} & \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s.t.} & \sum_i \lambda_i y_i = 0 \\ & \lambda \geq 0 \end{array}$$

其中 λ_i 为样本点 \mathbf{x}_i 的 Lagrange 乘子。KKT 条件定理指出，无效约束所对应的 Lagrange 乘子一定为 0，其几何意义为非支持向量 \mathbf{x}_i （对应于原问题中的约束 $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 > 0$ ）的 λ_i 一定为 0，这样分类规则就仅由少数的支持向量 \mathbf{x}_i （对应于 $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 = 0$ ）所决定，

而与其他样本无关。“支持向量机”这一名称由此而来。

针对 SVM 二次规划的求解，SMO (Sequential Minimal Optimization, 序列最小最优化) 是一种采用“分而治之”思想的快速算法，其不断地把原二次规划问题分解为只有两个变量（可达到的最小规模）的二次规划子问题，并对子问题用解析的方法进行求解，直到所有变量满足 KKT 条件为止。

10.1.4 澄清混淆：二次规划、二次收敛、二阶收敛

二次规划

二次规划 (Quadratic Programming) 是最简单的约束非线性规划，目标函数 (1) 是二次实函数 (如， x 为一维变量时， $f(x) = ax^2 + bx$ ； x 为 n 维变量时， $f(x) = \frac{1}{2}x^T Hx + c^T x$ ， H 为 n 阶对称矩阵)，约束函数 (2) 是线性的。二次规划简单易于求解，且一些非线性规划可转化为求解一系列二次规划问题。

二次收敛

二次收敛 (二次终止性) 是指一个算法用于正定二次型函数 (对于 $\forall x \neq 0$ ， $f(x) = x^T Hx > 0$ ， H 为 n 阶对称矩阵) 时，在有限步 (如 n 步) 内可达到它的极小点。

二阶收敛

二阶收敛与二次收敛根本就不是一回事。设 $\lim_{k \rightarrow \infty} x_k = x^*$ ，若有实数 $p \geq 1$ ，使

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} = c > 0$$

则称迭代方法为 p 阶方法。特别地， $p=1$ 时叫线性收敛； $p=2$ 时叫二阶收敛、平方收敛；对于 $p>1$ ，统称为超线性收敛。 p 的大小反映了收敛的快慢，越大收敛速度越快。 c 称为收敛因子、收敛比，当 p 相同时， c 越小则收敛速度越快。

10.2 最优化根基之单变量“一维搜索”

在后面 10.3 节中的众多的迭代极小化算法中，具有一个共同的特点，就是得到 (多维变量) 点 $x_k \in \mathbb{R}^n$ 后，需要按某种规则确定一个单位方向 d_k ($\|d_k\|=1$)，再从 x_k 出发，沿方向 d_k 在直线上求目标的极小点，从而得到 x_k 的后继点 $x_{k+1} = x_k + \lambda_k d_k$ ， λ_k 为步长。重复以上做法，直到求得问题的解。本节所讨论的求目标函数在直线上的极小点，称为一维搜索 (Line Search, 线搜索)，即单变量函数的极小化问题。

一维搜索方法归纳起来分为两大类。

- 试探法 (也被称为区间收缩法)，按某种方式找试探点，通过一系列试探点来确定极小点。典型的方法有：加步探索法 (进退法)、黄金分割搜索法、斐波那契搜索法。

- **函数逼近法**（也被称为插值法），用某种较简单的曲线逼近原本复杂的函数曲线，通过求逼近函数的极小点来估计目标函数的极小点。典型的方法有：**牛顿法**、**割线法**、**抛物线法**。

本节所讨论的一维搜索方法针对单维变量 $x \in \mathbf{R}^1$ ，因此在本节中不用考虑方向 d_k ，只需考虑步长 λ_k 即可： $x_{k+1} = x_k + \lambda_k$ 。此外，本节针对单谷函数（或叫单峰函数，Unimodal Function），如图 10-3 所示。

10.2.1 初始搜索区域的加步探索法（进退法）

我们首先介绍第一个大类：**试探法**，**不要求导计算**，因此属于**直接方法**，能适用于不可微的情况。其中，下面两节（10.2.2 节、10.2.3 节）介绍的两种方法（黄金分割搜索法、斐波那契搜索法）都要事先给定一个**包含极小点的单峰**初始搜索区域，本节中的**加步探索法（或称进退法、划界法 Bracket）**能解决这个问题。实际上，本方法本身也是一种区间试探法，主要思路就是从一点出发，按一定的步长，试图确定出函数值呈现“高一低一高”单峰状态的 3 点。首先从一个方向去找，若不成功，就退回来，再沿相反方向寻找，若方向正确，则加大步长进行探索，最终找到这 3 点为止。

具体地说，就是从初始点 x_0 （ $k=0$ ）开始，初始步长 $\lambda_0 > 0$ 。如果

$$f(x_k + \lambda_k) < f(x_k)$$

则下一步从新点 $x_k + \lambda_k$ 出发，加大步长 $\lambda_{k+1} = t\lambda_k$ （ $t > 1$ ）， $k := k+1$ ，再向前搜索，而如果

$$f(x_k + \lambda_k) > f(x_k)$$

若此时 $k=0$ ，则下一步仍以 x_0 为出发点，沿反方向搜索（ $\lambda_1 = -\lambda_0$ ）；否则就停止迭代，这样便得到一个包含极小点的初始搜索区间（点 x_k 与点 x_{k-2} 组成的区间）。

10.2.2 黄金分割搜索法（Golden Section Search）

黄金分割法适用于任何单峰函数求极小值问题。

首先介绍一下**黄金分割比**（也称为**中外比**）：把一条线段分割为两部分，使其中一部分与全长之比等于另一部分与这部分之比，即 $\frac{r}{1} = \frac{1-r}{r}$ ，其比值 r 是一个无理数 $r = \frac{\sqrt{5}-1}{2} \approx 0.618$ 。

在建筑学中，常使用黄金比例来达到美学效果，例如，代表了全希腊建筑艺术最高水平的帕特农神庙，其立面高与宽的比例为 19:31，接近黄金比例。

如图 10-5 所示，求函数在区间 $[a, b]$ 上的极小点，可通过不断缩小搜索区间来获得，即用一个含有极小值的子区间来代替最开始的区间 $[a, b]$ 。

具体地，我们在 $[a, b]$ 内取两点 c, d ，使得 $a < c < d < b$ 。我们选取的两个内点不是随意选定的，而是特别选定的，即人为地使 $[a, c]$ 与 $[d, b]$ 对称：

$$b-d = c-a \quad (3)$$

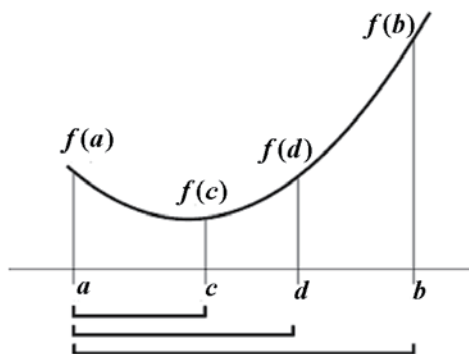


图 10-5 黄金分割搜索法

1. 如果 $f(c) < f(d)$ ，则最小点出现在 $[a, d]$ 上，因此 $[a, d]$ 成为下一次的搜索区间。
2. 如果 $f(c) > f(d)$ ，则 $[c, b]$ 成为下一次的搜索区间。

假如确定了 $[a, d]$ 是新的搜索区间，我们并不希望在 $[a, d]$ 上重新找两个新的点使之满足 (3) 式，而是利用已经找到的 c 点，再另找一个点即可，这样可大大节省工作量。也即，本次 c 点在区间 $[a, d]$ 的位置，就相当于上次 d 点在区间 $[a, b]$ 的位置：

$$\frac{d-a}{b-a} = \frac{c-a}{d-a} = r \quad (4)$$

其中 $r \in (0.5, 1)$ ，以保证 $c < d$ 。同时，我们希望 r 值在每一个子区间内均保持不变！

将公式 (3) 代入 (4)，可得： $r^2 + r - 1 = 0 \Rightarrow r = \frac{-1 \pm \sqrt{5}}{2}$

由于 $r \in (0.5, 1)$ ，故： $r = \frac{-1 + \sqrt{5}}{2} \approx 0.618$

事实证明，黄金比例搜索方法性能非常好，可达到**线性收敛**；相比之下，如果是随意放置的话，若运气不好就可能导致收敛很慢。注意 0.618 是黄金分割比 $\frac{\sqrt{5}-1}{2}$ 的一个近似值，试点最大个数为 10 才有意义（即最多只能做 9 次迭代）；如果取 0.618 033 988 7，则试点个数可增加到 15 个（即可做 14 次迭代）。

10.2.3 斐波那契 (Fibonacci) 搜索法

在一维搜索中，Fibonacci 搜索与黄金比例搜索相似，都是用分割区间的方法来求极小值。Fibonacci 搜索算法与 Fibonacci 数列有关。Fibonacci 数列用如下式子表达：

$$F_0 = 0, F_1 = 1, F_k = F_{k-1} + F_{k-2}$$

即，第 1 个数为 0，第 2 个数为 1，后面的每个数都是前两个数之和，例如 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, …

Fibonacci 搜索算法就是利用该数列进行区间的分割。与黄金比例搜索算法每次分割区间时使用固定的比例（0.618）不同，Fibonacci 搜索算法的区间缩短率是不固定的 F_{k-1}/F_k 。

在用分割方法求一维极小化问题时，Fibonacci 搜索算法是最优的策略，但缺点是需要事先确定搜索点的个数。相比之下，黄金比例搜索更简单，不需要事先知道计算次数，所以更常用。并且在实际中，为了能达到更快的收敛速度，通常会令黄金比例搜索配合使用逆抛物内插或其他超线性收敛技术，例如复杂的 Brent 算法，就是结合了黄金分割与逆抛物内插的可靠一维搜索算法。

Fibonacci 搜索算法是线性收敛的，它的极限形式正是黄金比例搜索，即：

$$\lim_{k \rightarrow \infty} \frac{F_{k-1}}{F_k} = \frac{\sqrt{5}-1}{2} \quad (\text{当 } k \geq 7 \text{ 时, } \approx 0.618)$$

10.2.4 牛顿法、抛物线法

前面介绍了试探法，我们接着介绍一维搜索中的函数逼近法，通过求导（要求导数存在）来进行解析计算，收敛速度更高。下面将介绍两种方法。

牛顿法

牛顿法的基本思想是，在极小点附近用二阶 Taylor（泰勒）展开多项式去近似代替原本复杂的目标函数 $f(x)$ ，进而求出极小点的估计值。

对于一维的单变量 $x \in \mathbf{R}^1$ ，我们需要求解： $\min f(x)$

二阶泰勒展开并舍去高阶项可得：

$$f(x) \approx f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2$$

因为在驻点处导数为 0，即 $f'(x_{k+1}) = f'(x_k) + f''(x_k)(x_{k+1} - x_k) = 0$ ，

$$\text{所以：} x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

牛顿法是以二阶速度收敛的。缺点是初始点的选择非常重要，要求充分接近极小点，否则就有可能不收敛。

抛物线法

抛物线法的基本想法是，在极小点（如图 10-6 所示红色的点 m ）附近，用二次三项式 $\varphi_k(x) = a_k + b_k x + c_k x^2$ （即一条抛物线，图中黑色的虚线）逼近目标函数 $f(x)$ （黑色实线）。

在3点 $x^{(1)} < x^{(2)} < x^{(3)}$ 处, $\varphi_k(x)$ 与 $f(x)$ 有相同的函数值, 并假设 $x^{(2)}$ 就是当前的极小点 x_k , 即3点中的最低点。

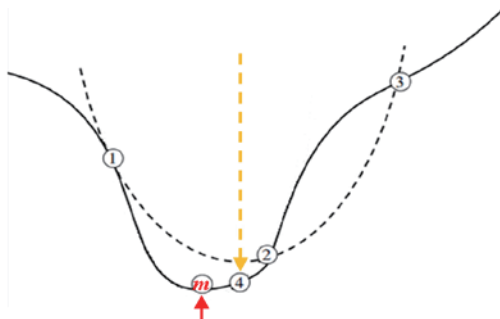


图 10-6 抛物线法

因为 $\varphi_k(x^{(1)}) = f(x^{(1)})$ 、 $\varphi_k(x^{(2)}) = f(x^{(2)})$ 、 $\varphi_k(x^{(3)}) = f(x^{(3)})$, 展开可得到3个方程, 联立可解出系数 a_k, b_k, c_k 。然后, 通过求导可得出 $\varphi_k(x)$ 的极小点 (驻点):

$$\varphi'_k(x^*) = b_k + 2c_k x^* = 0 \Rightarrow x^* = -\frac{b_k}{2c_k}$$

x^* (图中的第4点) 即为 x_{k+1} 。然后再组合左右两个邻点 (图中第1点和第2点), 构成新的3点 x_1, x_2, x_3 , 可看出, 相比之前的3点已将搜索区域缩小了。

10.2.5 不精确线搜索的 Armijo–Goldstein 准则及 Wolfe–Powell 准则

本节实际上讨论的是 10.3 节中多变量无约束优化的收敛性, 之所以提前放到本节, 是因为不精确一维搜索属于一维搜索的范畴。同时在本质上, 多变量优化一旦确定了搜索方向 d_k , 剩下的工作就是求解步长 λ_k 了, 这正是一维搜索的范畴, 本节实际上要讨论的就是如何确定步长 λ_k 的可接受区间。

我们首先解释一下什么是精确线搜索。如果最终求得的点 x_{k+1} 使得目标函数 $f(x)$ 达到极小值, 则称这样的一维搜索为**精确一维搜索**或**最优一维搜索**。而如果 x_{k+1} 只是使得 $f(x)$ 得到了可接受的下降量, 即下降量 $f(x_k) - f(x_{k+1})$ 是用户可接受的, 则称这样的一维搜索为**不精确一维搜索** (或近似一维搜索, 或可接受一维搜索)。在实际计算中, 精确一维搜索方法需要花费很大的工作量, 而且使用很多最优化方法, 如 10.3 节中的牛顿法和拟牛顿法, 其收敛速度并不依赖于精确一维搜索过程。因此, 我们只要保证目标函数 $f(x)$ 在每一步都有满意的下降即可, 因此更偏向于花费计算量较少的不精确一维搜索。

为了能让算法收敛, 需要遵循一些准则。我们首先提前解释一下 10.3 节中多变量不精确线搜索算法收敛的一般条件, 也即搜索方向 d_k 满足什么条件时算法才能收敛, 这是理解接下来两个准则的基础。

具体地, 每次优化后, 函数值必须是下降的, 即 $f(x_{k+1}) < f(x_k)$ 。我们将目标函数一阶泰勒展开并忽略掉高阶项, 可得:

由此可推出： $\lambda_k \mathbf{g}_k^T \mathbf{d}_k < 0 \Rightarrow \mathbf{g}_k^T \mathbf{d}_k < 0$ （其中步长 $\lambda_k > 0$ ， \mathbf{g}_k 为梯度）

$$\cos \theta_k = \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\| \|\mathbf{d}_k\|} > 0$$
$$0 \leq \theta_k \leq \frac{\pi}{2} - \mu, \quad \mu \in (0, \frac{\pi}{2})$$

Armijo-Goldstein 准则

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \lambda_k \rho \mathbf{g}_k^T \mathbf{d}_k \quad (5)$$

其中, $0 < \rho < \frac{1}{2}$, $\lambda_k > 0$

390

在图中，步长 λ_k 的取值区间为 $(0, a)$ 。可以看出，式 (5) 是为避免所选择的 λ_k 太靠近取值区间的端点 a ，这样就防止了目标函数 $f(\mathbf{x}_k + \lambda_k \mathbf{d}_k)$ 下降至太小。式 (6) 是为了避免步长 λ_k 太小，否则式 (5) 和式 (6) 的两条线就会重合在一起。

满足式 (5) 和式 (6) 的步长 λ_k 称为**可接受**步长因子，它的取值区间为 $[b, c]$ ，称为可接受区间。实际上，大多数情况下这个区间是能够将极小值点 k 包含进去的，本例中故意没有包含进去是为了引出下文。

Wolfe-Powell 准则

从图 10-7 中可以看到，Armijo-Goldstein 准则有可能把步长 λ_k 的极小值排除在可接受区间外面。因此，Wolfe-Powell 准则给出了一个更简单的条件来代替式 (6)：

$$\mathbf{g}_{k+1}^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k, \quad \sigma \in (\rho, 1) \quad (7)$$

其几何解释是在可接受点处切线的斜率大于或等于初始斜率的 σ 倍。从图中可以看出，其可接受区间为 $[e, c]$ 。一般而言， σ 值越小，线搜索越精确，但工作量越大。而不精确搜索不要过小的 σ ，通常可取 $\rho = 0.1, \sigma = 0.4$ 。

Wolfe-Powell 准则到这里还没有结束！在某些书中，你会看到用另一个所谓的“更强的条件”来代替 (7) 式，即：

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq -\sigma \mathbf{g}_k^T \mathbf{d}_k \quad (8)$$

这样就把可接受区间限制在了图中的 $[e, d]$ 范围之内，即将目标函数限定在图中加粗了的曲线（绿色）内。

10.3 多变量的无约束优化

10.3.1 最速下降法（Steepest Descent，梯度下降法 Gradient Descent）

考虑无约束问题：

$$\min f(\mathbf{x})$$

其中 $\mathbf{x} \in \mathbf{R}^n$ 为 n 维输入向量，目标函数 $f(\mathbf{x})$ 具有一阶连续偏导数。

人们总希望从某一点出发，选择一个使得目标函数值下降最快的方向，以便尽快达到极小点。如果采用**负梯度方向**作为下降方向，这种方法就是**最速下降法（Steepest Descent）**，也被称为**梯度下降法（Gradient Descent）**。

假设当前点所在的位置为 \mathbf{x}_k ，所对应的目标函数值为 $f(\mathbf{x}_k)$ ，那么如何使目标函数值进一步下降呢？我们可以将 \mathbf{x}_k 沿着某个单位向量方向 \mathbf{d}_k （ $\|\mathbf{d}_k\|=1$ ）移动步长 $\lambda_k > 0$ ，也即移动到了 $\mathbf{x}_k + \lambda_k \mathbf{d}_k$ 处，这时的目标函数值为：

$$f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) = f(\mathbf{x}_k) + \lambda_k \mathbf{g}_k^T \mathbf{d}_k + o(\lambda_k) \quad (9)$$

其中 $\mathbf{g}_k^T = \nabla f(\mathbf{x}_k)$ 为目标函数在 \mathbf{x}_k 这一点的梯度（一个向量）， $o(\lambda_k)$ 为 λ_k 的高阶无穷小。

如果你一时理解不了公式（9）的话，可将 \mathbf{x} 简化到一维的情况，就可看出该式实际上是个泰勒展开式：

$$f(x+h) = f(x) + f'(x)h + o(h)$$

其中， h 对应于 $\lambda_k \mathbf{d}_k$ ， $f'(x)$ 对应于 $\mathbf{g}_k^T = \nabla f(\mathbf{x}_k)$ 。这下是不是清楚了很多？

在公式（9）中，高阶无穷小可以忽略为 0，因此要使之取得极小值，则应使 $\mathbf{g}_k^T \mathbf{d}_k$ 最小。这实际上是两个向量的点积（数量积）。令向量 \mathbf{d}_k 与负梯度 $-\mathbf{g}_k$ 的夹角为 θ ，则 $\mathbf{g}_k^T \mathbf{d}_k = -\|\mathbf{g}_k\| \|\mathbf{d}_k\| \cos \theta = -\|\mathbf{g}_k\| \cos \theta$ 。因此 θ 为 0 时，该式取得最小值 $-\|\mathbf{g}_k\|$ ，也即方向 \mathbf{d}_k 取 $-\mathbf{g}_k$ 时，目标函数值下降得最快，这就是负梯度方向为“最速下降”方向的由来。

最速下降法的收敛性：对初始点的选择要求不高，一般的目标函数是整体收敛的（所谓整体收敛，是指不会非要在某些点附近的范围内，才会有好的收敛性）。

最速下降法的收敛速度：可以很快地从初始点达到极小点附近，至少**线性收敛**。但在**接近极小点时**，一般目标函数呈现为二次函数，会出现锯齿（Zigzagging）现象，且越靠近极小点步长越小，**收敛越来越慢**。所以适宜与其他方法结合用于**早期搜索**，比如前期用最速下降法，接近极小点时改用牛顿法。（最速下降法实际上是“名不副实”的，收敛非常缓慢，只能说是局部而非全局最快，而下面的牛顿法能够二阶收敛、收敛速度要更快。）

10.3.2 牛顿法（Newton）

上面的最速下降法只用到了梯度信息，即目标函数的一阶导数信息，本质上用线性函数逼近目标函数。而牛顿法则进一步用到了二阶导数信息。

将目标函数 $f(\mathbf{x})$ 在点 \mathbf{x}_k 处进行泰勒展开，只取二阶导数以及前面的项（更高阶被忽略掉），可得：

$$f(\mathbf{x}) \approx f(\mathbf{x}_k) + \mathbf{g}_k^T (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_k)^T \mathbf{H}_k (\mathbf{x} - \mathbf{x}_k) \quad (10)$$

其中 $\mathbf{g}_k^T = \nabla f(\mathbf{x}_k)$ 为目标函数在 \mathbf{x}_k 这一点的梯度（一个向量）； $\mathbf{H}_k = \nabla^2 f(\mathbf{x}_k)$ 为目标函数在 \mathbf{x}_k 这一点的**二阶导数矩阵**（即 Hesse/Hessian 矩阵，中文名海森、海赛、黑塞、海色）。

为了让 $f(\mathbf{x})$ 在 \mathbf{x}_k 取得极小值，则在该点处的一阶导数为 0，即对（10）式求导为 0：

$$\mathbf{g}_k + \mathbf{H}_k (\mathbf{x} - \mathbf{x}_k) = 0$$

当 \mathbf{H}_k 的逆矩阵存在，也即 \mathbf{H}_k 为非奇异矩阵时，上式两边都左乘逆矩阵 \mathbf{H}_k^{-1} ，得到：

$$\mathbf{H}_k^{-1} \mathbf{g}_k + (\mathbf{x} - \mathbf{x}_k) = 0 \quad \Rightarrow \quad \mathbf{x} = \mathbf{x}_k - \mathbf{H}_k^{-1} \mathbf{g}_k$$

由上式可知，为了取得极小值，则应将 \mathbf{x}_k 沿着 $-\mathbf{H}_k^{-1}\mathbf{g}_k$ 这个向量（含方向和大小）移动。

牛顿法的收敛速度：二阶收敛。因此，它比最速下降法要快。对于二次正定函数，只需一次迭代即可得到最优解。特别是在极小点附近，收敛性很好、速度快。注意：在牛顿法中不适合取恒定的步长，应采用某种一维搜索来确定步长，当步长收敛到 1 时，牛顿法才是二阶收敛的。

牛顿法的收敛性：对一般问题，只能保证局部收敛性，但不是整体收敛的（只有当初始点充分接近极小点时，才有很好的收敛性，否则不能保证收敛，甚至也不是下降方向），但通过一维搜索取非恒定步长可取得整体收敛。除了一维搜索，还可使用信赖域方法来保证总体收敛，而且信赖域进一步利用了 n 维二次模型，这样既具有牛顿法的快速局部收敛性，又具有理想的总体收敛性，还可以解决 Hesse 矩阵 \mathbf{H}_k 不正定和 \mathbf{x}_k 为鞍点等困难。

注意：Hesse 矩阵及其逆的求解计算量大，在迭代点处 Hesse 矩阵的逆也可能根本就不存在（即 Hesse 矩阵奇异）。

10.3.3 拟牛顿法（Quasi-Newton）：DFP 和 BFGS 方法

在牛顿法中，每一次要得到新的搜索方向的时候，都需要计算 Hesse 矩阵（二阶导数矩阵）。在自变量维数非常大的时候，这个计算工作是非常耗时的，因此，拟牛顿法的诞生就有意义了：它采用了一定的方法来构造与 Hesse 矩阵相似的 **正定矩阵**，而无须计算 Hesse 矩阵。实际上，**拟牛顿法属于一种共轭方向法**（见 10.3.4 节的介绍），如果初始矩阵取作单位阵，该方法就是共轭梯度法。

拟牛顿法与牛顿法的迭代过程一样，并构造与 Hesse 矩阵相似的正定矩阵，使用了目标函数的梯度（一阶导数）信息和两个点的“位移”（ $\mathbf{x}_{k+1} - \mathbf{x}_k$ ）来实现。有人会说，是不是用 Hesse 矩阵的近似矩阵来代替 Hesse 矩阵，会导致求解效果变差呢？事实上，效果反而通常会变好。因为在远离极小值点处，Hesse 矩阵一般不能保证正定，使得目标函数值不降反升；而拟牛顿法却可以使目标函数值沿下降方向走下去。

将目标函数 $f(\mathbf{x})$ 在点 \mathbf{x}_{k+1} 处进行泰勒展开，只取二阶导数以及前面的项（更高阶被忽略掉），可得：

$$f(\mathbf{x}) \approx f(\mathbf{x}_{k+1}) + \mathbf{g}_{k+1}^T (\mathbf{x} - \mathbf{x}_{k+1}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_{k+1})^T \mathbf{H}_{k+1} (\mathbf{x} - \mathbf{x}_{k+1})$$

两边对 \mathbf{x} 求导：

$$\nabla f(\mathbf{x}) = \mathbf{g}_{k+1}^T + \mathbf{H}_{k+1} (\mathbf{x} - \mathbf{x}_{k+1})$$

当 $\mathbf{x} = \mathbf{x}_k$ 时，有：

$$\mathbf{H}_{k+1}^{-1} (\mathbf{g}_{k+1}^T - \mathbf{g}_k^T) = \mathbf{A}_{k+1} (\mathbf{g}_{k+1}^T - \mathbf{g}_k^T) = \mathbf{x}_{k+1} - \mathbf{x}_k \quad (11)$$

上式中，我们令 $\mathbf{A}_{k+1} = \mathbf{H}_{k+1}^{-1}$ 。公式（11）就是拟牛顿方程，其中的矩阵 \mathbf{A}_{k+1} ，就是 Hesse 的逆矩阵的一个近似矩阵。在迭代过程中生成的矩阵序列 $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2 \dots$ 中（由于每一次迭

代尺度矩阵 \mathbf{A} 总是变化的，故方法也叫变尺度方法），对于某个矩阵 \mathbf{A}_{k+1} ，是由前一个矩阵 \mathbf{A}_k 修正得到的。修正方法有很多种，这里先说 DFP (Davidon-Fletcher-Powell) 方法的修正方法。

令： $\mathbf{q}_k = \mathbf{g}_{k+1}^T - \mathbf{g}_k^T$ ， $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ ，我们希望 \mathbf{A}_{k+1} 在 \mathbf{A}_k 的基础上加一个修正来得到：

$$\mathbf{A}_{k+1} = \mathbf{A}_k + \mathbf{B}_k \quad (12)$$

给定 \mathbf{B}_k 的一种形式：

$$\mathbf{B}_k = m\mathbf{v}\mathbf{v}^T + n\mathbf{w}\mathbf{w}^T \quad (13)$$

其中 m 和 n 均为实数， \mathbf{v} 和 \mathbf{w} 均为 n 维向量。

将 (12)，(13) 代入 (11) 可得：

$$\mathbf{v} = \mathbf{s}_k, \quad \mathbf{w} = \mathbf{A}_k \mathbf{q}_k, \quad m = 1/(\mathbf{v}^T \mathbf{q}_k), \quad n = 1/(\mathbf{w}^T \mathbf{q}_k)$$

下面再给一种拟牛顿法——BFGS (Broyden-Fletcher-Goldfarb-Shanno) 方法。

$$\mathbf{A}_{k+1} = \mathbf{A}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{q}_k} - \frac{\mathbf{A}_k \mathbf{q}_k \mathbf{q}_k^T \mathbf{A}_k}{\mathbf{q}_k^T \mathbf{A}_k \mathbf{q}_k} + (\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{A}_k (\mathbf{g}_{k+1} - \mathbf{g}_k) \mathbf{w} \mathbf{w}^T \quad (14)$$

$$\text{其中：} \mathbf{w} = \frac{\mathbf{s}_k}{\mathbf{s}_k^T \mathbf{q}_k} - \frac{\mathbf{A}_k \mathbf{q}_k}{\mathbf{q}_k^T \mathbf{A}_k \mathbf{q}_k}$$

公式 (14) 中右边的前面 3 项就是 DFP 方法，最右边的第 4 项就是 BFGS 比 DFP 多出来的部分。

BFGS 方法是迄今为止最好的拟牛顿方法，能够整体收敛。与 DFP 一样，只要初始矩阵对称正定，则 BFGS 修正公式所产生的矩阵也是对称正定的，具有二次收敛性。而且矩阵不易变为奇异，因此 BFGS 比 DFP 具有更好的数值稳定性。

拟牛顿法的缺点是所需内存较大，对于大型问题，可能会遇到内存不足的问题。因此有一种方法叫 Limited-Memory BFGS，简称 L-BFGS 或 LM-BFGS（这里的“LM”与 Levenberg-Marquard 算法没有关系），比 BFGS 算法占用的内存小。

10.3.4 共轭方向法 (Conjugate Direction)

最速下降法有锯齿现象，收敛速度慢；而牛顿法需要计算 Hesse 矩阵而计算量大。共轭方向法收敛速度界于两者之间，具有二次收敛性，对于正定 n 维二次函数最多经 n 次一维搜索即可收敛。由于一般目标函数在极小点附近呈现为二次函数，既然共轭方向法对于二次函数比较有效，所以对原目标函数也有较好效果。

共轭方向法仅需利用一阶导数信息，但克服了最速下降法收敛慢的缺点，又避免了存储和计算牛顿法所需的二阶导数信息。共轭方向法是在研究对称正定二次函数的基础上提出来的。

$$\text{令：} \mathbf{p}_m \mathbf{Q} \mathbf{p}_n = 0$$

则称两个向量（搜索方向） $\mathbf{p}_m, \mathbf{p}_n$ 为对称正定矩阵 \mathbf{Q} 的共轭向量。特别地，当 \mathbf{Q} 为单位矩

阵 E 时, 有 $p_m p_n = 0$, 因此共轭是正交的推广。

共轭方向法的基本思想是把一个 n 维问题转化为 n 个一维问题。对于正定 n 维二次函数, 从任意点 x_0 出发, 沿任意下降方向 p_0 做直线搜索得到 x_1 。再从 x_1 出发, 沿与 p_0 共轭的方向 p_1 做直线搜索, 经过 n 次迭代必定收敛于正定二次函数的极小值。将这个思路转化为公式形式:

$$\begin{cases} \min_{\lambda} f(x_k + \lambda p_k) = f(x_k + \lambda_k p_k) \\ x_{k+1} = x_k + \lambda_k p_k \end{cases}$$

为确定最优步长 λ_k , 只需对 $f(x_k)$ 求导为 0 即可: $\frac{df(x_{k+1})}{d\lambda} = 0$

现在的问题是如何产生一组关于对称正定矩阵 Q 共轭的向量? 这里介绍一种叫作 Gram-Schmidt 的方法。

取线性无关的向量组 v_0, v_1, \dots, v_{n-1} , 例如取 n 个坐标轴的单位向量。初始化 $p_0 = v_0$, 则:

$$p_{k+1} = v_{k+1} - \sum_{j=0}^k \frac{p_j^T Q v_{k+1}}{p_j^T Q p_j} p_j$$

共轭方向法属于效果好而又实用的方法。理论与实践证明, 将二次收敛方法用于非二次的目标函数, 亦有很好的效果, 但迭代次数不一定保证有限次, 即对非二次 n 维目标函数经 n 步共轭方向一维搜索不一定就能达到极小点。在这种情况下, 为了找到极小点, 可用泰勒级数将该函数在极小点附近展开, 略去高于二次的项之后即可得该函数的二次近似。实际上, 很多函数都可以用二次函数很好地近似, 甚至在离极小点不是很近的点也是这样。

10.3.5 共轭梯度法 (Conjugate Gradient)

共轭梯度法是最著名的一种共轭方向法。初始共轭向量 p_0 由初始迭代点 x_0 处的负梯度 $-g_0$ 来给出, 使得最速下降方向具有共轭性; 而以后的 p_k 由当前迭代点的负梯度与上一个共轭向量的线性组合来确定:

$$p_{k+1} = -g_{k+1} + \lambda_k p_k \quad (15)$$

$$\lambda_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad (16)$$

其中公式 (16) 仅用到梯度信息产生了 n 个搜索方向, 此公式称为 Fletcher-Reeves 公式 (F-R 公式), 通常称为 F-R 共轭梯度法。

共轭梯度法对于正定 n 维二次函数必在 n 次以内收敛。对于非二次函数的优化问题, 迭代次数不止 n 次, 但共轭方向只有 n 个。更重要地, 经过 n 步迭代之后, 产生的新方向不再有共轭性。所以在实际运用中, 也有很多修正方向的策略。其中一种策略是: 经过 n 步迭代之后, 取负梯度方向作为新的方向。

共轭梯度法的收敛性：共轭梯度法仅比最速下降法稍微复杂一点，却具有二次收敛性，比最速下降法要好得多。对于正定二次函数，具有二次收敛性。而且，不用求矩阵的逆，所需内存量较小，因此适用于求解变量较多的大规模问题。

10.3.6 Powell 直接法

下面介绍**不需要对输入变量求导数**的方法，这类方法一般称为**直接方法**。直接方法与对输入变量求导数的方法相比，一般来说收敛得比较慢。但因为不需要对输入变量求导数，所以迭代比较简单，编程也比较容易。根据数值计算的经验，对于变量不多的问题，能够收到较好的效果。

Powell 方法是一种直接搜索法，该方法本质上是共轭方向法。整个计算过程分成若干个迭代，每轮迭代由 $n+1$ 次一维搜索组成，即先依次沿着已知的 n 个**线性无关方向**（如取各个维度的单位方向）进行一维搜索，得到一个“最好”点，然后沿本阶段的初点与该“最好”点连线方向进行一维搜索，求得这一阶段最终的“最好”点。然后，再用最终的搜索方向取代前 n 个方向之一，开始下一阶段的迭代。具体计算步骤如下：

1. 给定初始点 \mathbf{x}_0 ， n 个线性无关的方向 $\mathbf{d}_1^{(1)}, \mathbf{d}_1^{(2)}, \dots, \mathbf{d}_1^{(n)}$ ，以及允许误差 $\varepsilon > 0$ 。
2. 第 $k=1, 2, \dots$ 个迭代阶段，令 $\mathbf{x}_k^{(0)} = \mathbf{x}_{k-1}$ ，从 $\mathbf{x}_k^{(0)}$ 出发，依次沿方向 $\mathbf{d}_k^{(1)}, \mathbf{d}_k^{(2)}, \dots, \mathbf{d}_k^{(n)}$ 进行一维搜索，依次得到最优点： $\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)}, \dots, \mathbf{x}_k^{(n)}$ 。

再从 $\mathbf{x}_k^{(n)}$ 出发，沿着方向 $\mathbf{d}_k^{(n+1)} = \mathbf{x}_k^{(n)} - \mathbf{x}_k^{(0)}$ 做一维搜索，得到点 \mathbf{x}_k 。

3. 若 $\|\mathbf{x}_k - \mathbf{x}_{k-1}\| < \varepsilon$ ，则停止计算，得到 \mathbf{x}_k ；否则，令

$$\mathbf{d}_{k+1}^j = \mathbf{d}_k^{j+1}, \quad j=1, \dots, n, \quad \text{返回步骤 (2) 迭代。}$$

Powell 直接法具有二次收敛性。但如果在某轮迭代中， n 个搜索方向线性相关，则有可能不收敛。

10.4 最优化根基之“信赖域”

在本节中，将讲述一下信赖域方法与一维搜索的区别、联系，以及信赖域方法的数学思想、实现过程。

最优化的目标是找到极小值点。在这个过程中，我们可以从某一个点 \mathbf{x}_k 开始，先确定一个搜索方向 \mathbf{d}_k ，在这个方向上做一维搜索（Line Search）。找到此方向上的可接受点之后，调整搜索方向，然后继续在新的方向上进行一维搜索。即通过“调整搜索方向→进行一维搜索→调整搜索方向”的迭代步骤，直到我们认为目标函数已经收敛到了极小值点。具体地，在一维搜索中，从 \mathbf{x}_k 点移动到下一个点的过程，可以描述为： $\mathbf{x}_k + \lambda_k \mathbf{d}_k$ ，此处 $\lambda_k \mathbf{d}_k$ 就是在 \mathbf{d}_k 方向上的位移，可以记为 \mathbf{s}_k 。

信赖域和一维搜索同为最优化方法的基础方法，但它**不像一维搜索那样先求搜索方向 \mathbf{d}_k 再求步长 λ_k** ，而是**每次根据“某种原则”直接确定位移 \mathbf{s}_k** ，即直接在一个区域内试图找到一个好

的试探点。该区域称为信赖域，通常是以当前迭代点为中心的一个小邻域。试探点往往要求是**原优化问题的某个近似问题**在信赖域的解，求出后判断它是否可以被接受为下一个迭代点。试探点的好坏还被用来决定如何调节信赖域，如果试探点（即根据“某种原则”确定的位移）能使目标函数值充分下降，则保持不变或扩大信赖域；若不能使目标函数值充分下降，则缩小信赖域。如此迭代下去，直到收敛。

那么，到底根据什么原则来直接确定位移呢？可**利用相对比较简单的二次模型去近似目标函数 $f(\mathbf{x})$** ，再用二次模型计算出位移 \mathbf{s}_k 。根据位移 \mathbf{s}_k 可以确定下一点 $\mathbf{x}_k + \mathbf{s}_k$ ，从而可以计算出目标函数的下降量（下降是最优化的目标），再根据下降量来决定扩大信赖域或缩小信赖域。

那么，该如何判定要扩大还是缩小信赖域呢？为了说明这个问题，必须先描述信赖域方法的数学模型：

$$\begin{cases} \min m(\mathbf{s}_k) = f(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T \mathbf{H}_k \mathbf{s}_k & (17) \\ s.t. \quad \|\mathbf{s}_k\| \leq h_k & (18) \end{cases}$$

（17）式就是我们用于近似目标函数的二次模型，其自变量为 \mathbf{s}_k ，也就是我们要求的位移。 \mathbf{g}_k 为梯度， \mathbf{H}_k 为 Hesse 矩阵。如果 Hesse 矩阵不好计算，可以利用“有限差分”来近似，或者用拟牛顿方法来构造 Hesse 矩阵的近似矩阵。

（18）式中的 h_k 是第 k 次迭代的信赖域上界（或称为信赖域半径），即位移要在信赖域上界范围内。此外，第 2 个式子中的范数没有指定是什么范数，例如，是 2-范数还是 ∞ -范数。

现在回到了上面的问题：该如何判定要扩大还是缩小信赖域？通过衡量二次模型与目标函数的近似程度，可以做出判定：

第 k 次迭代的二次模型拟合的下降量为： $\Delta m_k = f(\mathbf{x}_k) - m(\mathbf{s}_k)$

但第 k 次迭代的实际下降量为： $\Delta f_k = f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{s}_k)$

定义比值： $r_k = \Delta f_k / \Delta m_k$ ，这个比值可以用于衡量二次模型与目标函数的近似程度，显然 r 值越接近 1 越好。

由此，我们给出信赖域方法的基本迭代步骤。

1. 从初始点 \mathbf{x}_0 ，初始信赖域半径 $h_0 = \|\mathbf{g}_0\|$ 开始迭代。
2. 到第 k 步时，计算 \mathbf{g}_k 和 \mathbf{H}_k 。
3. 求解信赖域模型，求出位移 \mathbf{s}_k ，并计算 r_k 。
4. 若 $r_k \leq 0.25$ ，说明步子迈得太大了，应缩小信赖域半径，令 $h_{k+1} = \|\mathbf{s}_k\|/4$ 。
5. 若 $r_k \geq 0.75$ 且 $\|\mathbf{s}_k\| = h_k$ ，说明这一步已经迈到了信赖域半径的边缘，并且步子有点小，可以尝试扩大信赖域半径，令 $h_{k+1} = 2h_k$ 。
6. 若 $0.25 < r_k < 0.75$ ，说明这一步迈出去之后，处于“可信赖”和“不可信赖”之间，可

以维持当前的信赖域半径，令 $h_{k+1} = h_k$ 。

7. 若 $r_k \leq 0$ ，说明函数值是向着上升而非下降的趋势变化的（与最优化的目标相反），这说明这一步迈得错得“离谱”了，这时不应该走到下一点，而应“原地踏步”，即 $\mathbf{x}_{k+1} = \mathbf{x}_k$ ，并且和上面 $r_k \leq 0.25$ 的情况一样缩小信赖域。反之，在 $r_k > 0$ 的情况下，都可以走到下一点，即 $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ 。

10.4.1 Levenberg-Marquardt (L-M) 方法

Levenberg-Marquardt 方法是最重要的一种信赖域方法。在公式 (18) 中，当信赖域模型中的范数 $\|\mathbf{s}_k\| \leq h_k$ 取 2-范数时（即 $\|\mathbf{s}\|_2 \leq h_k$ ），就得到了 Levenberg-Marquardt 方法（简称 L-M 方法）的数学模型：

$$\begin{cases} \min m(\mathbf{s}_k) = f(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T \mathbf{H}_k \mathbf{s}_k \\ \text{s.t. } \|\mathbf{s}_k\|_2 \leq h_k \end{cases}$$

L-M 算法是介于牛顿法与梯度下降法之间的一种非线性优化方法，对于过参数化问题不敏感，能有效处理冗余参数问题，使代价函数陷入局部极小值的机会大大减小，这些特性使得 L-M 算法在计算机视觉等领域得到广泛应用。如视觉计算中的运动与结构估计问题、捆绑调整问题，通常都是过参数化，且变量成千上万，L-M 算法性能优越，结合稀疏矩阵的特点可把计算复杂度大大降低。L-M 算法本质上是在迭代过程中把非线性最小二乘问题化为多个线性最小二乘问题在信赖域上的求解（如果没有信赖域约束，就是 Gauss-Newton 法）。

10.4.2 详解 L-M 方法的求解过程与步骤

除了前面介绍的基本迭代步骤，L-M 方法还有很多种实现方式，下面就介绍一种常见的解法。

我们考虑一个参数拟合问题。已知函数关系 $f: f(\mathbf{p}) = \mathbf{y}$ ，其中 \mathbf{p} 是待求解的 n 维参数向量； \mathbf{y} 是已知的 m 维观测向量（而不是一维的标量）。实际上，也同时给定了一组已知的对应 \mathbf{y} 的 m 维向量 \mathbf{x} ，但因为此时 $f(\mathbf{p}, \mathbf{x}) = f(\mathbf{p})$ ，故在式中不显示。计算步骤为：

1. 取初始点 \mathbf{p}_0 ，终止控制常数 δ ，计算 $\varepsilon_0 = \|\mathbf{y} - f(\mathbf{p}_0)\|$ ，令 $k := 0$ ， $\lambda_0 = 10^{-3}$ 。

2. 对于 \mathbf{p}_k 的下一个点 $\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{s}_k$ ，其函数值 $f(\mathbf{p}_{k+1}) = f(\mathbf{p}_k + \mathbf{s}_k)$ 的一阶泰勒展开为：

$f(\mathbf{p}_k + \mathbf{s}_k) \approx f(\mathbf{p}_k) + \mathbf{J}_k \mathbf{s}_k$ ，其中 \mathbf{J}_k 为 Jacobi 矩阵 $\frac{\partial f}{\partial \mathbf{p}}|_{\mathbf{p}=\mathbf{p}_k}$ 。每次迭代，我们都希望找到一个位移向量 \mathbf{s}_k ，使得

$\|\mathbf{y} - f(\mathbf{p}_k + \mathbf{s}_k)\| \approx \|\mathbf{y} - f(\mathbf{p}_k) - \mathbf{J}_k \mathbf{s}_k\| = \|\boldsymbol{\varepsilon}_k - \mathbf{J}_k \mathbf{s}_k\|$ 最小，即 $\boldsymbol{\varepsilon}_k = \mathbf{J}_k \mathbf{s}_k$ ，得到正规化方程组： $(\mathbf{J}_k^T \mathbf{J}_k) \mathbf{s}_k = \mathbf{J}_k^T \boldsymbol{\varepsilon}_k$ 。

我们将这个方程组略加修改，得到增量正规化方程组：

$$(\lambda_k \mathbf{I} + \mathbf{J}_k^T \mathbf{J}_k) \mathbf{s}_k = \mathbf{J}_k^T \boldsymbol{\varepsilon}_k \quad (19)$$

在 L-M 算法中，每次迭代是寻找一个合适的阻尼因子 λ_k ，当 λ_k 很小时，算法就变成了 Gauss-Newton 法； λ_k 很大时，算法则为梯度下降法。相比于 Gauss-Newton 法中的矩阵 $\mathbf{J}_k^T \mathbf{J}_k$ ，增加一个正定对角矩阵可改变原矩阵的特征值，变成条件数较好的对称正定矩阵。

3. 求解增量正规化方程组 (19) 得到位移向量 \mathbf{s}_k 。

(3a) 如果 $\|\mathbf{y} - f(\mathbf{p}_k + \mathbf{s}_k)\| < \varepsilon_k$ ，则令 $\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{s}_k$ ，如果 $\|\mathbf{s}_k\| < \delta$ ，停止迭代，输出结果。否则令 $\lambda_{k+1} = \lambda_k / 10$ ，置 $k := k + 1$ ，转到第 2) 步。

(3b) 如果 $\|\mathbf{y} - f(\mathbf{p}_k + \mathbf{s}_k)\| \geq \varepsilon_k$ ，则令 $\lambda_{k+1} = 10\lambda_k$ ，重新解正规化方程组得到 \mathbf{s}_k ，返回步骤 3a)。

10.5 最小二乘问题的求解

我们经常会遇到如下形式的最小二乘（最小平方）问题，可看作是无约束极小化的特殊情况。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。

$$\min_{\mathbf{x} \in R^n} f(\mathbf{x}) = \sum_{i=1}^m [r_i(\mathbf{x})]^2 = \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) \quad (20)$$

可看出 $f(\mathbf{x})$ 是 m 个子函数 $r_i(\mathbf{x})$ 的平方和。

10.5.1 线性最小二乘问题的求解（正规化方法、QR 分解、SVD 分解）

如果 $r_i(\mathbf{x})$ 都为线性函数（如 $x_1 - 2$ ， $3x_1 + 2x_2 - 5x_3 + 1$ 等），则问题 (20) 是线性最小二乘问题。对 \mathbf{x} 向量的每个变量求偏导，可得到一个线性方程组：

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (21)$$

其中 \mathbf{A} 是个 $m \times n$ 的矩阵， \mathbf{x} 为 n 维向量， \mathbf{b} 为 m 维向量。一般 $m > n$ ，约束个数大于未知量个数，即问题是超定的 (Overdetermined)，否则问题就是欠定的 (Underdetermined)。

假设公式 (21) 中的矩阵 \mathbf{A} 是满秩的，即 $\text{rank}(\mathbf{A}) = n$ ，则该问题称为满秩最小二乘问题。

正规化方法：

将 (21) 转化为求解正规化方程组（或法方程组 Normal Equations）：

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$$

通过正规化方法的求解步骤为：

1. 对 $(\mathbf{A}^T \mathbf{A})$ 进行 Cholesky 分解 $(\mathbf{A}^T \mathbf{A}) = \mathbf{G}\mathbf{G}^T$ ，

2. 求解方程组 $\mathbf{G}\mathbf{y} = (\mathbf{A}^T \mathbf{b})$ 得到 \mathbf{y} ，进一步求解方程组 $\mathbf{G}^T \mathbf{x} = \mathbf{y}$ 得到 \mathbf{x} 。

注意第 1) 步中，除了 Cholesky 分解 (Cholesky Decomposition)，还有更一般的 LU 分解

(LU Factorization)。实际上，LU 分解可用于一般矩阵，Cholesky 分解则针对对称正定矩阵，是 LU 分解的特例。比起一般的 LU 分解，计算 Cholesky 分解更为快捷，并具有更高的数值稳定性。

QR 分解方法：

通过 QR 分解的求解步骤为：

1. $A = QR$ ， Q 是个 $m \times n$ 的正交矩阵；
2. 取 Q 的前 n 列构成一个矩阵 Q_1 ，即 $Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$ ；取 R 的前 n 行构成一个矩阵 R_1 ；
3. 求解方程组 $R_1 x = Q_1^T b$ 得到 $x = R_1^{-1} Q_1^T b$ 。

QR 分解方法比正规化方法精确，也比较稳定，特别是当 A 的条件数较大（病态）时。但当 $m \gg n$ 时，运算量大约是正规化方法的两倍； $m = n$ 时，几乎相同。

奇异值分解（SVD）方法：

首先定义矩阵 A 的奇异值分解： $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ ，则 A 的广义逆矩阵（也被称为 Penrose-Moore 广义逆） $A^+ = V \begin{pmatrix} \Sigma^{-1} \\ 0 \end{pmatrix} U^T$ 。最后，通过 $x = A^+ b$ 得到 x 。

SVD 分解速度最慢，但最可靠。

10.5.2 非线性最小二乘问题（Gauss-Newton 方法）

实际上，前面提到的 Levenberg-Marquardt 方法就可用于求解非线性最小二乘问题。本节将介绍另一种称为 Gauss-Newton 的方法。

Gauss-Newton 方法基于前面提到的 Newton 方法的思路。将目标函数 $f(x)$ 在点 x_k 处进行泰勒展开，只取二阶导数以及前面的项（更高阶被忽略掉），可得：

$$f(x) = f(x_k) + g_k^T (x - x_k) + \frac{1}{2} (x - x_k)^T H_k (x - x_k)$$

在最小二乘形式中， $f(x)$ 的梯度 $g(x) = J(x)^T r(x)$ ，其中 $J(x)$ 是 $r(x)$ 的 Jacobi 矩阵； $f(x)$ 的 Hesse 矩阵为 $H(x) = J(x)^T J(x) + S(x)$ ，其中 $S(x)$ 为 Hesse 矩阵中的二阶项，通常难以计算因此将其忽略掉，因此 $H(x) \approx J(x)^T J(x)$ 。

跟前面提到的 Newton 方法相似，有：

$$x = x_k - H_k^{-1} g_k = x_k - (J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k)$$

由于 Newton 方法是局部二阶收敛的，因此 Gauss-Newton 方法的成功依赖于所忽略的二阶项 $S(x)$ 在 $H(x)$ 中的重要性，根据情况的不同可得到二阶收敛、线性收敛或不收敛。在实际中，我们常加上一维搜索策略，称为阻尼 Gauss-Newton 法，使得能够整体收敛。此外，

$H(\mathbf{x}) \approx J(\mathbf{x})^T J(\mathbf{x})$ 应该是可逆，非奇异的；如果奇异的话，可考虑采用信赖域策略，这就变成了 Levenberg-Marquardt 方法（见 10.4.1 节）。

对于线性最小二乘问题，Gauss-Newton 方法可一步达到极小点。

10.6 约束优化问题的求解

一般而言，求解带有约束条件的问题比无约束优化要困难：在每次迭代时，不仅要使目标函数值有所下降，而且要使迭代点落在可行域内。我们可将约束问题转化为一个或一系列无约束问题的求解，将复杂的问题变为多个较简单的子问题等。

二次规划是非线性规划的特殊情形，其目标函数是二次实函数，约束是线性的。二次规划比较简单，易于求解，且一些非线性规划可转化为求解一系列二次规划问题。

10.6.1 等式约束的拉格朗日乘子法 (Lagrange Multiplier)

比如我们要求解下面有等式约束的优化问题：

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = 0 \end{aligned}$$

我们定义 **Lagrange 函数**： $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda h(\mathbf{x})$ ，其中 $\lambda \neq 0$ 称为 **Lagrange 乘子**。为了获得 Lagrange 函数的极小值，分别对 \mathbf{x} 、 λ 求偏导数，可得：

$$\begin{cases} \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \nabla_{\mathbf{x}} h(\mathbf{x}) = 0 \\ h(\mathbf{x}) = 0 \end{cases}$$

注意：Lagrange 函数 $L(\mathbf{x}, \lambda)$ 的极小化等价于原目标函数 $f(\mathbf{x})$ 的极小化，这是因为满足上面这个方程组的极值点 \mathbf{x}^* 有 $h(\mathbf{x}^*) = 0$ ，所以 $L(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*)$ 。联立求解上面这个方程组便可得到 \mathbf{x}^* ，代入 $f(\mathbf{x})$ 即可得到目标函数的极小值。

10.6.2 不等式约束的 KKT (KT) 条件

对于如下含有不等式约束的优化问题，如何求取最优值呢？

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h(\mathbf{x}) = 0 \\ g(\mathbf{x}) \geq 0 \end{aligned} \quad (22)$$

常用的方法是 **KKT (Karush-Kuhn-Tucker) 条件**，也被称为 **KT (Kuhn-Tucker) 条件**，是拉格朗日乘子法的推广。同样地，把目标函数、等式约束和不等式约束全部写为一个式子 $L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda \times h(\mathbf{x}) + \mu \times g(\mathbf{x})$ ，这被称为 **广义拉格朗日函数 (Generalized Lagrange Function)**。KKT 条件指的是最优值必须满足以下条件：

1. $L(\mathbf{x}, \lambda, \mu)$ 对 \mathbf{x} 求偏导为 0 : $\nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \times \nabla_{\mathbf{x}} h(\mathbf{x}) + \mu \times \nabla_{\mathbf{x}} g(\mathbf{x}) = 0$
2. $h(\mathbf{x}) = 0, \lambda \neq 0$
3. $\mu \times g(\mathbf{x}) = 0, \mu \geq 0$

求取这 3 个等式之后就能得到最优值。其中第 3 个式子非常有趣，因为 $g(\mathbf{x}) \geq 0$ ，如果要满足这个等式，必须 $\mu = 0$ 或者 $g(\mathbf{x}) = 0$ ，这被称为**互补松弛条件 (Complementary Slackness Condition)**。这是 SVM 的很多重要性质的来源，如支持向量的概念。

注意：Lagrange 乘子法和 KKT 条件所求结果是最优解要满足的**必要条件**，但不一定是最优解。**只有在凸函数的情况下，才能保证也是充分条件，即保证求得的是最优解**，这是因为凸函数下的局部最优解就是全局最优解。

实际上，(22) 所定义的原目标函数极小化问题等同于广义拉格朗日函数的**极大极小问题** $\min_{\mathbf{x}} \max_{\lambda, \mu} L(\mathbf{x}, \lambda, \mu)$ ，即计算广义拉格朗日函数的**鞍点**。而根据**博弈论 (Game Theory)** 所推导出的**拉格朗日对偶性 (Lagrange Duality)**，它们的对偶问题为**极大极小问题** $\max_{\lambda, \mu} \min_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu)$ 。在凸规划情况下，若满足 KKT 的对偶互补条件，则原始问题和对偶问题的最优值相等。在 SVM 中，我们将原问题的极小值求解转化为 **Wolfe 对偶表示**的极大值求解。

10.6.3 惩罚函数法（外点法、内点法）

对于约束优化问题 (22)，我们可以构造惩罚函数（罚函数，Penalty Function） $p(\mathbf{x}, M)$ 来将其转变为无约束优化问题：

$$p(\mathbf{x}, M) = f(\mathbf{x}) + M \{ [\min(0, g(\mathbf{x}))]^2 + [h(\mathbf{x})]^2 \}$$

其中 $\min(0, g(\mathbf{x}))$ 表示取 0 和 $g(\mathbf{x})$ 两者之间的最小值。右式中的第 2 项称为**惩罚项**； $M > 0$ 称为**罚因子**，为一个充分大的正数。

可以看出，惩罚函数只对不满足约束条件的点实行惩罚。比如若 $g(\mathbf{x}_k) < 0$ 或者 $h(\mathbf{x}_k) \neq 0$ ，则平方后再乘以一个充分大的正数 M ，会使得此时的惩罚函数 $p(\mathbf{x}_k, M)$ 变得非常大，离极小化的目标变得很遥远。因此在最优化过程中，惩罚函数会时刻保证满足约束条件。

罚因子 M 不能选得过小，否则惩罚函数的极小点就会远离约束问题的最优解； M 也不能选得过大，否则极小点位于一个十分狭长的深谷之中，搜索方向稍有偏离就会导致相当大的误差。因此，可由小到大地选取一个因子序列，被称为**序列**无约束技术（SUMT, Sequential Unconstrained Minimization Technique）。随着 M 值的增加，惩罚函数中惩罚项所起的作用越来越大，即对点远离**可行域**（满足所有约束方程的点组成的集合）的惩罚越来越重，这就迫使惩罚函数的极小点与可行域的距离越来越近。当 M 趋于正无穷大时，迭代点就从可行域外部趋于极小点，**外点法**正是由此得名。

如果惩罚函数在可行域边界上取值为无穷，则称为**内点法**，其只适用于不等式约束，其惩罚函数定义为：

$$p(\mathbf{x}, r) = f(\mathbf{x}) + r \frac{1}{g(\mathbf{x})}$$

因此, 内点法只取可行域内部的点, 不包括可行域边界上的点, 即 $g(\mathbf{x}) \neq 0$ 。与外点法不同, 内点法要求整个迭代过程始终在可行域内部进行。通过上式右边第 2 项可以看出, 惩罚函数在可行域边界上形成了一堵无穷高的“障碍墙”, 所以内点惩罚函数也被称为**障碍罚函数**。

小结比较一下外点法和内点法。**外点法**在整个空间内进行优化, 初始点可以任意给定; 但中间结果不是可行解, 不能作为近似最优解, 只有迭代到最后才能得到符合要求的可行解。**内点法**总是在可行域内进行, 每个中间结果都是可行解, 可作为近似解; 但初始点选取较困难, 且只适用于不等式约束。

惩罚函数法的优点是方法简单, 使用方便, 并能用来求解导数不存在的问题。但收敛速度慢、计算量大, 每次迭代都需求解一个无约束优化问题。此外, 求解过程要求罚因子无限增大, 可能导致 Hesse 矩阵严重病态。解决办法是, 可将罚项加入到拉格朗日乘子法中进行改进, 定义增广 Lagrange 函数 (乘子罚函数), 以克服这些缺点。

10.7 最短路径与动态规划 (Dynamic Programming)

如图 10-8 所示, 在**图论 (Graph Theory)** 中一个抽象的图 (Graph) 包括一些节点和连接它们的边, 如果每条边是有长度的, 即**权重**, 则这个图就被称为**加权图 (Weighted Graph)**。**最短路径 (Shortest Path)** 问题是图论中的经典问题, 比如要找到起点 S 与终点 T 之间的最短距离。最笨的方法是穷举法, 把所有可能的路线都测量一遍, 比如在本图中从 S 到 T 共有 $3 \times 3 \times 2 \times 1 = 18$ 条不同的线路, 如果节点数更多的话, 乘积将是个非常大的数字。

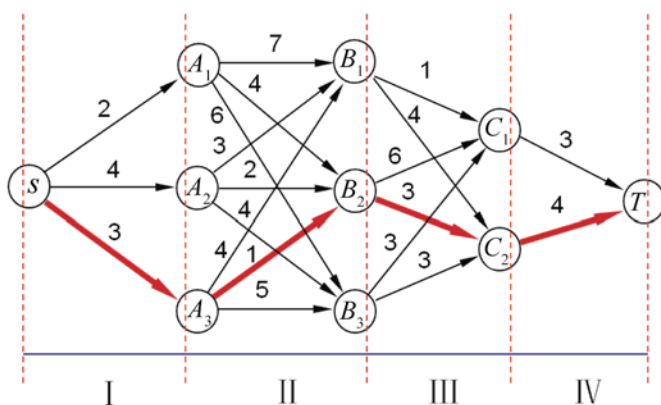


图 10-8 动态规划用于求解最短路径

我们可用**动态规划 (DP, Dynamic Programming)** 的方法来快速实现最短路径导航。基本思想是分而治之, 将一个比较复杂的问题分解成**互相联系 (不是互相独立)**的一系列**同一类型 (即重叠性质)**的更容易求解的子问题 (阶段), 即变成一个**多阶段最优决策问题**。图 10-8 中, 我们

把从 S 到 T 的全过程分成了 I、II、III、IV 共 4 个阶段。这也是动态规划中“动态”一词的由来，即人为地引进“时间”因素，分成时段。

动态规划的另一个特点是**最优子结构**，即**最优解包含着其子问题的最优解**。以图 10-8 为例，我们不妨反过来思考这个最短路径问题，假设已经找到了 $S \rightarrow A_3 \rightarrow B_2 \rightarrow C_2 \rightarrow T$ 这一条最短路径。那么在第 III 阶段，**最短路径上的节点 B_2 到终点 T 的路线 $B_2 \rightarrow C_2 \rightarrow T$ ，也是从节点 B_2 到终点 T 一切可能路径中的最短路径！**这就是 Bellman 提出的著名的**最优性原理 (Principle of Optimality)**。因为如果不是的话，它就不是最终的最短路径。正因为对每个子问题都考虑到最优效果，于是就排除了大量的中间非最优的方案组合，使计算量比穷举法大大减少。

利用这个性质，我们可以**从最后一个阶段 (IV) 开始**，由终点 T 向起点 S 逐阶段**递推**，寻求各点到终点 T 的最短路径。当递推到起点 S 时，便是全过程的最短路径。这种由后向前逆向递推的方法正是动态规划常用的**逆序法 (后向法)**。在此过程中，我们**保存**已解决的子问题的最优路径，在需要时直接调用即可，不必重新计算，这就可以避免大量重复计算，即“**用 (存储) 空间换 (计算) 时间**”。比如逆向递推到第 I 阶段，要计算起点 S 经过 A_1 到终点 T 的最短距离，直接将“ SA 的边长”和“第 II 阶段已求得的 A_1 到终点 T 的最短距离”相加即可。

此外还可看出，动态规划问题的局限是：状态**必须满足无后效性**（即**马尔科夫性**，Markov）。所谓的**无后效性**是指：下一时刻的状态只与当前状态有关，而和当前状态之前的状态无关，即当前的状态是对以往决策的总结。

动态规划可用于**背包问题、旅行商问题、生产库存计划问题、资源分配问题、设备更新问题**。

对于最短路径问题，除了**动态规划 (如 Floyd 算法)**，可求解全源最短路径，即任意一对节点的最短路径)，还有 **Dijkstra 算法**， **A^* (A-Star、A 星) 算法**。这几种算法思路是相似的，不同点仅在于“推进点的选取”和“算法终止条件”。其中 Dijkstra 算法是一种求单源最短路径的算法，即从一个点开始到所有其他点的最短路，且边的权重不能为负数。 A^* 算法可视为 Dijkstra 算法的一个扩展，可利用已知信息来估计某一点到目标点的距离，以减小最短路径的搜索范围，提高了效率，缺点是空间需求很大，为指数级别。

10.8 “偶然中的必然”——概率与贝叶斯 (Bayes)

“概率论只不过是把常识用数学公式表达了出来。”

——拉普拉斯

10.8.1 先验概率、似然函数、后验概率、贝叶斯公式

联合概率 $p(x, y)$ 的乘法公式： $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$

(如果随机变量 x, y 是**独立**的，则 $p(x, y) = p(x)p(y)$)

由乘法公式可得**条件概率公式：** $p(x|y) = \frac{p(x, y)}{p(y)}$ ， $p(y|x) = \frac{p(x, y)}{p(x)}$

全概率公式： $p(x) = \sum_{m=1}^M p(x|y_m)p(y_m)$ ，其中 $\sum_{m=1}^M p(y_m) = 1$

($\sum_{m=1}^M p(y_m) = 1$ ，则 $p(x) = \sum_{m=1}^M p(x, y_m)$ 可轻易推导出上式)

$$\text{贝叶斯公式: } p(y_m|x) = \frac{p(x, y_m)}{p(x)} = \frac{p(x|y_m)p(y_m)}{\sum_{m=1}^M p(y_m)p(x|y_m)}$$

又名后验概率公式、逆概率公式：后验概率 $p(y_m|x)$ = 似然函数 $p(x|y_m)$ × 先验概率 $p(y_m)$ / 证据因子 $p(x)$ 。解释如下，假设我们根据“手臂是否很长”这个随机变量 x （取值为“手臂很长”或“手臂不长”）的观测样本数据来分析远处一个生物是猩猩类别 y_1 还是人类类别 y_2 （假设总共只有这两种类别）。我们身处一个人迹罕至的深山老林里，且之前就有很多报道说这里有猩猩出没，所以无须观测样本数据就知道是猩猩的先验概率（Prior Probability） $p(y_1)$ 较大，比如根据历史数据估计有 $70\% = 0.7$ 。接着，我们得到了 x 的观测样本数据：“手臂很长”——而猩猩 y_1 类别表现为这种特征的类条件概率，或者说这种“可能性”即似然（Likelihood） $p(x|y_1)$ 相比于人类表现为“手臂很长”的似然 $p(x|y_2)$ 较大。所以经这次观测之后加强了我们的判断：是一只猩猩的后验概率（Posterior Probability） $p(y_1|x)$ 变得比先验概率更大，超过了之前的 70% ！反之，如果观测发现这个生物的手臂不长，而猩猩类别 y_1 表现为“手臂不长”的似然较小，则会减弱我们的判断，猩猩的后验概率将小于 70% 。因此，后验概率包含了先验信息以及观测样本数据提供的后验信息，对先验概率进行了修正，更接近真实情况。此外，证据因子（Evidence，也被称为归一化常数） $p(x)$ 可仅看成一个权值因子，以保证各类别的后验概率总和为 1 从而满足概率条件。

如果我们的目标仅仅是要对所属类别做出一个判别：是“猩猩”还是“人类”，则无须去计算后验概率的具体数值，只需计算哪个类别的后验概率更大即可。假设猩猩和人类出现的先验概率相等， $p(y_1) = p(y_2) = 0.5$ ，则此时类别的判定完全取决于似然 $p(x|y_1)$ 和 $p(x|y_2)$ 的大小。因此，似然函数（Likelihood，“可能性”）的重要性不是它的具体取值，而是当参数（如类别参数） y 变化时，函数到底变小还是变大，以便反过来对参数进行估计求解（估计出是 y_1 还是 y_2 ）。

10.8.2 朴素（Naïve）贝叶斯分类

假设我们有 K 个（比如 1 000 个）样本点组成的样本集合 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_K\}$ ，每个样本点 $\mathbf{x} \in \mathbf{R}^N$ 是一个 N 维向量， $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}, \dots, x^{(N)})$ ；样本点的类别标签集合 $\mathbf{Y} = \{y_1, \dots, y_m, \dots, y_M\}$ ，即共有 M 种不同的类别标签。说直白点，样本点 \mathbf{x} 是有 N 个特征的向量，比如像第 6 章的 6.4.1 节那样，我们对人体或猩猩样本用“手臂长度”、“嘴巴突起的高度”、“门牙长度”这 3 个特征进行描述，此时样本点 $\mathbf{x} \in \mathbf{R}^3$ 就是一个三维的向量。注意本例中每个维度特征的取值是连续的，我们可分别将每一维特征离散化，各转为 S_n 种离散的取值 $\{x_1^{(n)}, \dots, x_s^{(n)}, \dots, x_{S_n}^{(n)}\}$ ，比如对将第 n 维的 $x^{(n)}$ 取值划分为 $S_n = 2$ 个分区，其中一个分区的数值表示“长的手臂” $x_1^{(n)}$ ，另一个分区的数值表示“短的手臂” $x_2^{(n)}$ 。此外，在这个例子中，我

们只有两个类别（人体、猩猩），因此此时 $M = 2$ ，即类别标签集合为 $\mathbf{Y} = \{y_1, y_2\}$ 。

分类问题要解决的事情就是：给定一个新的样本点 \mathbf{x} ，我们要估计它的类别标签 y （是人体类别 y_1 ，还是猩猩类别 y_2 ）。为了让计算机能完成这件事，你首先要训练它，比如事先给定 $K = 1000$ 个样本点，并手工指定每一个样本点的标签，比如 (\mathbf{x}_1, y_1) ， (\mathbf{x}_2, y_1) ， (\mathbf{x}_3, y_2) ， \dots ， (\mathbf{x}_{1000}, y_1) ，让计算机进行学习。那么，如何让计算机根据手工给定的样本进行训练学习，并对新的未知类别标签的样本点进行自动分类呢？本节中会介绍一种称为朴素贝叶斯（Naïve Bayes）的方法。

1. 首先，朴素贝叶斯方法根据样本点组成的训练数据集，去学习联合概率分布。比如，对于给定类别标签 y_m ，联合概率分布 $p(\mathbf{x}, y_m)$ 可利用先验概率分布 $p(y_m)$ 和类条件概率分布 $p(\mathbf{x} | y_m)$ 来求解：

$$p(\mathbf{x}, y_m) = p(y_m)p(\mathbf{x} | y_m)$$

为了求得类条件概率分布 $p(\mathbf{x} | y_m)$ ，朴素贝叶斯方法对各个维度的特征做了条件独立性的简化假设：

$$p(\mathbf{x} | y_m) = p(x^{(1)}, x^{(2)}, \dots, x^{(N)} | y_m) = \prod_{n=1}^N p(x^{(n)} | y_m)$$

条件独立意味着每个特征的分布都可以独立地被当作一维分布来估计，这样减轻了由于“维数灾难”带来的影响，当样本的特征个数增加时就不需要使样本规模呈指数增长。条件独立性是个较强的假设，非常简单、朴素，“too simple, sometimes naïve”，这一假设使算法变得简单，但有时会牺牲一定的分类准确率。

因此，在朴素贝叶斯方法中，学习就意味着需要对先验概率 $p(y_m)$ 和类条件概率 $p(x^{(n)} | y_m)$ 进行估计，以得到它们的值。下一节（10.8.3 节）将介绍 3 种估计方法。

2. 假设我们学习到了联合概率分布，下面我们进行分类，即对一个新的样本点 \mathbf{x}_* ，通过学习到的模型计算后验概率分布 $p(y_m | \mathbf{x}_*)$ ，并将后验概率最大的类别（比如 $p(y_1 | \mathbf{x}_*) > p(y_2 | \mathbf{x}_*)$ ）作为 \mathbf{x}_* 的类别标签（ y_1 ）。那么，如何计算后验概率分布 $p(y_m | \mathbf{x}_*)$ 呢？可利用贝叶斯公式：

$$p(y_m | \mathbf{x}_*) = \frac{p(y_m)p(\mathbf{x}_* | y_m)}{\sum_{m=1}^M p(y_m)p(\mathbf{x}_* | y_m)} = \frac{p(y_m) \prod_{n=1}^N p(x_*^{(n)} | y_m)}{\sum_{m=1}^M p(y_m)p(\mathbf{x}_* | y_m)}$$

上式就是朴素贝叶斯分类的基本公式。对于给定的样本点 \mathbf{x}_* ，上式的分母 $\sum_{m=1}^M p(y_m)p(\mathbf{x}_* | y_m)$ 是个常数，故略去。我们只比较分子即可：

$$p(y_m) \prod_{n=1}^N p(x_*^{(n)} | y_m)$$

即，如果 $p(y_1) \prod_{n=1}^N p(x_*^{(n)} | y_1) > p(y_2) \prod_{n=1}^N p(x_*^{(n)} | y_2)$ ，则可认为这个样本点属于类别 y_1 。

理论上，**贝叶斯分类器在统计模式识别中被称为最优分类器**。但在实际应用中，先验概率和类条件概率密度一般情况下很难知道。因此，朴素贝叶斯方法做了简化，假设输入特征向量的各个维度都是条件独立的，这就又有点太简化了。因为现实中各个特征属性间往往并不是条件独立的，即存在概率依赖关系，这时模型可考虑变成一种有向无环图（DAG）：**贝叶斯网络**（又称**贝叶斯信念网络**、**信念网络 Belief Network**）。

小结一下。**朴素贝叶斯属于生成式方法（Generative Approach）**，一开始需要根据样本数据去学习联合概率分布函数 $p(\mathbf{x}, y)$ ，涉及类条件概率函数 $p(\mathbf{x} | y_m)$ 的估计。然后在分类时，对于一个新样本 \mathbf{x}_* ，通过贝叶斯公式求解后验概率 $p(y | \mathbf{x}_*)$ ，并选取使之最大的那个类别标签 y_m 即可。学习联合概率分布实际上需要估计一个**能够符合样本数据分布的概率模型**，于是训练样本被认为正是由这个模型所**生成**的。当有大量训练样本时，可学习到接近真实的模型。生成式模型是所有变量的全概率模型，因此能够用于生成任意变量的值。当存在隐变量时，生成式方法依然适用。生成式模型能够反映同类数据本身的相似度，但它不关心分类边界在哪，因此需要转化成后验概率以进行分类。其他的生成式方法还有**高斯混合模型（GMM, Gaussian Mixture Model）**、**隐马尔可夫模型（HMM, Hidden Markov Model）**、**深度信念网（DBN, Deep Belief Networks）**等。

然而，一般来说，概率模型（比如数据符合哪种概率分布，以及具体的分布参数）是比较难估计的，特别是在样本数据不多的情况下。而第6章介绍的**K近邻、决策树、感知机、SVM、AdaBoost、逻辑斯谛回归、条件随机场（CRF, Conditional Random Field）**等则属于**判别式方法（Discriminative Approach）**，直接寻找不同类别之间的分类边界（如直接采用判别决策函数 $y = f(\mathbf{x}_*)$ 或概率函数 $p(y | \mathbf{x}_*)$ ），反映的是异类数据之间的差异，准确率也更高。判别式方法不考虑样本的产生模型，不关心这些样本的概率分布，只要能把不同类别的数据很好地分开就行，还可很方便地对数据先进行预处理以降低分类的难度。然而，判别式模型通常需要解决凸优化问题。对于隐变量，判别式方法也不能处理。

由生成式模型可以得到判别式模型，但由判别式模型得不到生成式模型。

10.8.3 最大似然估计、最大后验概率估计、贝叶斯估计

本节我们给出3种方法，估计训练样本所属的统计模型的参数 θ 。比如工厂产品 \mathbf{x} 是合格品 y_1 还是废品 y_2 符合一个统计分布，我们可以通过抽样（采样）质检一批产品样本，来估计其中合格品类别 y_1 的概率 $\theta = p(y_1)$ 这一参数（即合格率）。又比如，我们假定属于某一类别的多个训练样本点符合同一正态分布，需要估计该正态分布的具体参数，如均值、方差。

最大似然估计（MLE）

设随机变量 \mathbf{x} 的 K 个训练样本组成的集合 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$ 符合同一个概率分布

$p(x|\theta)$ ，其中 θ 是未知的模型参数。则这 K 个样本的联合分布概率密度函数为：

$$L(\mathbf{X}|\theta) = L(x_1, x_2, \dots, x_K|\theta) = \prod_{k=1}^K p(x_k|\theta)$$

L 是一个关于变量 θ 的函数， x_1, x_2, \dots, x_K 是已知的。由于 θ 不是随机变量，而是个“未知但固定”的量，所以函数 L 不是关于随机变量的函数，严格来讲不能称为概率（Probability），通常称为“似然（Likelihood）”，即似然函数。最大似然（ML，Maximum Likelihood）参数估计就是求参数 θ^* ，以使得似然函数 L 达到最大值：

$$\theta^* = \arg \max_{\theta} L(x_1, x_2, \dots, x_K|\theta) = \arg \max_{\theta} \prod_{k=1}^K p(x_k|\theta)$$

此外由于对数函数的单调性，为了简化求解（将相乘变成相加），有对数似然函数：

$$L(\theta) = \ln \prod_{k=1}^K p(x_k|\theta) = \sum_{k=1}^K \ln p(x_k|\theta)$$

为了使得对数似然函数达到最大值，很简单，求导为 0 即可：

$$\frac{\partial \sum_{k=1}^K \ln p(x_k|\theta)}{\partial \theta} = \sum_{k=1}^K \frac{1}{p(x_k|\theta)} \frac{\partial p(x_k|\theta)}{\partial \theta} = 0$$

可以看出，最大似然估计并没有用到先验概率 $p(\theta)$ ，没有用到任何贝叶斯模型。

我们回到 10.8.2 节的朴素贝叶斯方法。首先求解类条件概率 $p(x^{(n)}|y_m)$ 的最大似然估计。本例中在预处理时，我们将 $x^{(n)}$ 原始连续值通过划分变成了离散值。如果是原始的连续值，则应该怎么办呢？通常假定其服从高斯分布（也称正态分布） $N(\mu, \sigma^2)$ ，可通过最大似然估计先求解出属于 y_m 类别的 $x^{(n)}|_{y_m}$ 的均值 $\mu^{(n)}|_{y_m}$ 和方差 $\sigma^{(n)}|_{y_m}$ 这两个中间参数，进一步求解便可得到类条件概率。

下面给出离散情况下， $x^{(n)}$ 取第 s 种离散值 $x_s^{(n)}$ （比如“手臂长度”是“短的”），条件概率 $p(x_s^{(n)}|y_m)$ 的最大似然估计：

$$p(x_s^{(n)}|y_m) = \frac{\sum_{k=1}^K I_{x_k}(x_s^{(n)}, y_m)}{\sum_{k=1}^K I_{x_k}(y_m)}, \quad n=1, 2, \dots, N; \quad s=1, 2, \dots, S_n; \quad m=1, 2, \dots, M$$

其中 $I_{x_k}(y_m)$ 为指示函数（Indicator Function），为了统计样本集中类别标签是 y_m 的样本个数。比如训练样本共有 $K=1000$ 个点，若某个点 x_k 的类别标签是 y_m ，则 $I_{x_k}(y_m)$ 为 1；否则为 0，不参与个数的统计。

先验概率 $p(y_m)$ 的最大似然估计是：

$$p(y_m) = \frac{\sum_{k=1}^K I_{x_k}(y_m)}{K}, \quad m = 1, 2, \dots, M$$

最大后验概率 (MAP) 估计

在最大似然估计中，未知参数 θ 不认为是个随机变量，即是个“未知但固定”的量。而在最大后验概率估计中，未知参数 θ 被认为是个随机变量。则后验概率 $p(\theta | \mathbf{X})$ 为：

$$p(\theta | \mathbf{X}) = \frac{p(\theta)p(\mathbf{X} | \theta)}{p(\mathbf{X})}$$

最大后验概率 (MAP, Maximum a Posteriori Probability) 估计就是求参数 θ^* ，使得后验概率 $p(\theta^* | \mathbf{X})$ 达到最大值。很简单，求导为 0 即可：

$$\frac{\partial}{\partial \theta} (p(\theta)p(\mathbf{X} | \theta)) = 0$$

注意式中没有包括 $p(\mathbf{X})$ ，因为它相对于 θ 是独立的。最大似然法和最大后验概率估计的不同在于：后者加入了先验分布 $p(\theta)$ 。而假设 \mathbf{X} 服从均匀分布，即 $p(\theta)$ 对于所有的 θ 是一个常量，这两种估计就会得到同样的结果。但在一般的分布情况下，结果会不同。实际上，最大后验概率估计渐近地趋向于最大似然估计值。

贝叶斯估计

贝叶斯估计也把未知参数 θ 看作随机变量，是在 MAP 上做进一步拓展，此时不直接估计参数的值，而是允许参数 θ 服从一定的概率分布，具有已知的先验分布 $p(\theta)$ 。样本通过似然函数 $p(\mathbf{X} | \theta)$ 并利用贝叶斯公式将 θ 的先验分布 $p(\theta)$ 转化为后验分布 $p(\theta | \mathbf{X})$ ：

$$p(\theta | \mathbf{X}) = \frac{p(\theta)p(\mathbf{X} | \theta)}{p(\mathbf{X})} \propto p(\theta)p(\mathbf{X} | \theta)$$

后验分布 ($|$) 包含了关于 θ 的先验信息以及样本提供的后验信息，是通过观测后得到的，比先验分布 $p(\theta)$ 更接近真实情况。最后， θ 的贝叶斯估计 θ^* 是在给定样本集合 \mathbf{X} 下的条件期望：

$$\theta^* = E[\theta | \mathbf{X}] = \int_{\Theta} p(\theta | \mathbf{X}) \theta d\theta$$

同样，我们回到 10.8.2 节的朴素贝叶斯方法。用最大似然估计可能会出现所要估计的概率值为 0 的情况，而贝叶斯估计则不会。当 $x^{(n)}$ 取第 s 种离散值 $x_s^{(n)}$ （比如“手臂长度”是“短的”）时，其类条件概率 $p(x_s^{(n)} | y_m)$ 的最大似然估计是：

$$p(x_s^{(n)} | y_m) = \frac{\sum_{k=1}^K I_{x_k}(x_s^{(n)}, y_m) + \lambda}{\sum_{k=1}^K I_{x_k}(y_m) + S_n \lambda}, \quad n = 1, 2, \dots, N; \quad s = 1, 2, \dots, S_n; \quad m = 1, 2, \dots, M$$

其中 $\lambda \geq 0$ ，等价于在随机变量各个取值的频数上加一个正数 $\lambda > 0$ 。当 $\lambda = 0$ 时，就是最大似然估计。常取 $\lambda = 1$ ，此时称为**拉普拉斯平滑 (Laplace smoothing)**。

先验概率 $p(y_m)$ 的最大似然估计是：

$$p(y_m) = \frac{\sum_{k=1}^K I_{x_k}(y_m) + \lambda}{K + M \lambda}, \quad m = 1, 2, \dots, M$$

从以上 3 种估计方法可以看出，一般而言，从 MLE 到 MAP 再到贝叶斯估计，对参数的表示越来越精确，越来越能够反映基于样本的真实参数情况。

10.8.4 贝叶斯学派与频率学派之争论

在前面有关概率统计的介绍中，我们频繁地看到了**贝叶斯**这个词，有**贝叶斯公式**、**贝叶斯分类**、**贝叶斯估计**、**贝叶斯网络**等。实际上，贝叶斯学派起源于 18 世纪一个不很知名的新教牧师 Thomas Bayes（托马斯·贝叶斯，1702—1761 年）的一篇遗作，而且还是他的朋友在整理其个人数学遗稿时才发现的，本人并没有打算要发表出来。这个学说在接下来的 100 多年里受到强烈排斥和非议，乃至沉寂了很久。从 20 世纪二三十年代开始，贝叶斯理论才又重新不断发展，并由此在概率统计学领域出现了**贝叶斯学派 (Bayesians)** 和经典的**频率学派 (Frequentists)** 的争论。

贝叶斯学派的两个基本观点是：

1. 把待估参数 θ 看作随机变量，而频率学派则视 θ 为未知常数。
2. 待估参数 θ 在抽样观测前就具有先验概率分布，而频率学派认为任何模型都不存在先验。

实际上，比如你操作一台 3D 打印机所产生的废品率 θ ，长期来看确实不是固定的，会有波动（而且随着你操作经验的提升，3D 打印的废品率 θ 还会有下降的趋势），可看作一个随机变量。这个参数符合一个先验分布。抽样（采样）获得的样本数据会有各种各样的偏差，而一个合适的先验分布则可以排除这些随机噪声的干扰。

贝叶斯学派善于利用过去的先验知识和样本数据进行逻辑归纳推理，而频率学派仅仅利用样本数据。因此贝叶斯推论中前一次得到的后验概率分布可作为后一次的先验概率，是一个随“时间”推移而“增量”认识的过程。此外，贝叶斯学派认为所有的参数都是随机变量，都有分布，因此可使用一些基于采样的方法使得我们更容易构建复杂模型，比如 MCMC（Markov Chain Monte Carlo，马尔科夫链的蒙特卡洛采样）方法提供了从后验分布直接采样的途径，为贝叶斯统计方法的实际应用带来了革命性的突破。贝叶斯学派的优势在于：对小样本很有效（因为很多事件是不可重复的，无法进行多次实验），也适合于高维的参数空间。此外，贝叶斯推断问题

相对简单，点估计、区间估计和假设检验全部可以由后验分布得到，尤其是计算机技术的发展 and MCMC 方法的出现，使得非共轭后验分布的使用和计算成为可能，而且它的理论架构天然符合人类渐进增量的认识规律。

频率学派最重要的观点就是不断重复（越多越好，趋近于无限），并且认为模型的参数是固定的：一个模型在无数次的抽样过后，所有的参数都应该是一样的。在频率学派看来，**概率**指的是相对**频率**，是真实世界的**客观**属性；而在贝叶斯学派看来，概率描述的是**主观信念**的程度，并不是频率——这种认知方式可让我们不仅能对随机变化产生的数据进行概率描述，还能对各个参数进行概率描述，即便它们是固定的常数。频率学派的优势是没有假设一个先验分布，因此更加客观，也更加无偏，因此在一些保守的领域（比如制药业、法律）比贝叶斯学派更受到信任，可构造长期稳定的性能（如置信区间）。

贝叶斯学派的困难在于“先验分布如何确定”。所有的先验在参数变换后都不可避免地带有主观性，显得不太客观。目前已有一些解决方法，如无信息先验分布、共轭先验分布、最大熵方法确定先验分布等。而频率学派用最大似然估计（MLE）则没有这个问题，但频率学派的困难在于如何利用前人已有经验和枢轴统计量的构造。

几十年来两个学派争论不休，都曾经相互断言对方的理论必将灭亡，但目前都还看不到迹象。而这期间两者的折中“经验贝叶斯”却发展起来了。经验贝叶斯与传统贝叶斯的不同是，它用数据来估计 MMLE（Marginal Maximum Likelihood Estimator）先验分布中的参数，因此为一些频率学派学者所接受。

值得一提的是，贝叶斯学派最常关心的是后验分布，频率学派最常关心的是似然函数。而我们知道，后验分布其实就是似然函数乘以先验分布再除以一个证据因子，因此两者的很多方法其实都是相通的。

总之，正如量子力学中“波动论”和“粒子论”的争论，两方最后被“波粒二象性（Wave-Particle Duality）”统一了起来，也许贝叶斯学派和频率学派的争论最后也会有一个更好的框架来进行统一。

全书到此已结束，中间章节历经了多个流派，从最开始的“**大局宏观派**”，到“**操作实战派**”，再到“**技术方法派**”、“**商业运作派**”，一直到最后的“**学院理论派**”，本书对 3D 打印与 3D 智能数字化的各个门派都已分别进行了详细论述。相信你终会在“全球第三次工业革命”这个荡气回肠的“江湖”中寻到自己的门派，确立自己的江湖地位，并乐在其中！

（全书完）

参考文献

- [1] 马颂德, 张正友著. 计算机视觉——计算理论与算法基础. 科学出版社, 2003
- [2] 吴怀宇. 3D 数字化与 3D 打印: 用“虚拟”再造“现实”(上篇). 中国科学报, 2013-4-10
- [3] 吴怀宇. 3D 数字化与 3D 打印: 转向“中国智造”的产业机遇(下篇). 中国科学报, 2013-4-17
- [4] 吴怀宇. 3D 打印: 智能数字化. 光明日报, 2013-09-17 (12)
- [5] 吴怀宇. 智能数字化与 3D 打印: “中国智造”推动“全球第三次工业革命”. 中国自动化学会通讯, 2013, (2): 38~45
- [6] 吴怀宇. 3D 打印把人工智能梦想照进现实. 北京科技报, 2013-7-1
- [7] 吴怀宇. 基于离散微分几何的数字几何处理研究 (PhD thesis). 中国科学院, 2008
- [8] Huai-Yu Wu, Chunhong Pan, Hongbin Zha, Qing Yang, Songde MA. Partwise Cross-Parameterization via Nonregular Convex Hull Domains. IEEE Transactions on Visualization And Computer Graphics, 2011
- [9] Huai-Yu Wu, Chunhong Pan, Qing Yang, Songde MA. Consistent Correspondence between Arbitrary Manifold Surfaces. ICCV, 2007
- [10] Huai-Yu Wu, Hongbin Zha. Robust Consistent Correspondence Between 3D Non-Rigid Shapes Based On ‘Dual Shape-DNA’. ICCV, 2011
- [11] Huai-Yu Wu, Hongbin Zha, Tao Luo, Xu-Lei Wang, Songde MA. Global and Local Isometry-Invariant Descriptor for 3D Shape Comparison and Partial Matching. CVPR, 2010
- [12] Huai-Yu Wu, Chunhong Pan, Hongbin Zha, Songde MA. Model Transduction for Triangle Meshes. Journal of Computer Science and Technology (JCST), 2010, 25(3): 584-595
- [13] (美) 彼得·马什 (Peter Marsh). 新工业革命. 中信出版社, 2013
- [14] (美) 杰里米·里夫金 (Jeremy Rifkin). 第三次工业革命. 中信出版社, 2012

- [15] (美) 克里斯·安德森 (Chris Anderson). 创客：新工业革命. 中信出版社, 2012
- [16] (美) 胡迪·利普森, 梅尔芭·库曼. 3D 打印：从想象到现实, 中信出版社, 2013
- [17] 吴福朝. 计算机视觉中的数学方法. 科学出版社, 2008
- [18] 王飞跃. 从社会计算到社会制造：一场即将来临的产业革命. 中国科学院院刊, 2012, 27 (6), 658~669
- [19] 周昆. 数字几何处理：理论与应用 (PhD thesis). 浙江大学, 2002
- [20] Shi-Min Hu et al. Internet visual media processing: a survey with graphics and vision applications. The Visual Computer, 2013, 29(5), 393-405
- [21] T.F.Cootes, G.J.Edwards, and C.J.Taylor. Active appearance models. IEEE TPAMI, 2001, 23, 681-685
- [22] Rosten Edward, Reid Porter, and Tom Drummond. FASTER and better: A machine learning approach to corner detection. IEEE TPAMI, 2010, 32: 105-119
- [23] Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. CVPR, 2001
- [24] Geoffrey E. Hinton, Ruslan R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. Science, 2006-7-28, 313(5786): 504-507
- [25] 余凯. 深度学习：推进人工智能的梦想. 程序员, 2013.6
- [26] Blanz, V., Vetter, T. A Morphable Model for the Synthesis of 3D Faces. SIGGRAPH, 1999
- [27] Olga Sorkine, et al. Laplacian Surface Editing. SGP, 2004
- [28] Yizhou Yu, Kun Zhou, Dong Xu, Xiaohan Shi, Hujun Bao, Baining Guo, Heung-Yeung Shum. Mesh Editing with Poisson-Based Gradient Field Manipulation. SIGGRAPH, 2004.
- [29] Jean-Sebastien Franco, Edmond Boyer. Exact Polyhedral Visual Hulls. BMVC, 2003
- [30] Vladimir G. Kim, Wilmot Li, Niloy Mitra, Stephen DiVerdi, and Thomas Funkhouser. Exploring Collections of 3D Models using Fuzzy Correspondences. SIGGRAPH, 2012
- [31] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007
- [32] Linjie Luo, Ilya Baran, Szymon Rusinkiewicz and Wojciech Matusik. Chopper: Partitioning Models into 3D-Printable Parts. SIGGRAPH Asia, 2012
- [33] Bruno Vallet, Bruno Lévy. Spectral Geometry Processing with Manifold Harmonics. Eurographics, 2008
- [34] Romain Prévost, Emily Whiting, Sylvain Lefebvre, Olga Sorkine-Hornung. Make It Stand: Balancing Shapes for 3D Fabrication. SIGGRAPH, 2013

- [35] Weiming Wang, Tuanfeng Y. Wang, Zhouwang Yang, Ligang Liu, et al. Cost-effective Printing of 3D Objects with Skin-Frame Structures. ACM Transactions on Graphics (Proc. SIGGRAPH Asia), 2013, 32(5), Article 20: 1-10
- [36] Takeo Igarashi, Satoshi Matsuoka, Hidehiko Tanaka. Teddy: A Sketching Interface for 3D Freeform Design. SIGGRAPH, 1999
- [37] Andrew Nealen, Takeo Igarashi, Olga Sorkine, Marc Alexa. FiberMesh: Designing Freeform Surfaces with 3D Curves. SIGGRAPH, 2007
- [38] Tao Chen, Zhe Zhu, Ariel Shamir, Shi-Min Hu, Daniel Cohen-Or. 3-Sweep: Extracting Editable Objects from a Single Photo. ACM Transactions on Graphics (TOG), SIGGRAPH Asia, 2013, 32(6)
- [39] Yoav Freund, Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997, 55(1):119-139
- [40] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror and Dani Lischinski. Data-Driven Enhancement of Facial Attractiveness. SIGGRAPH, 2008
- [41] Haoda Huang, Jinxiang Chai, Xin Tong, Hsiang-Tao Wu. Leveraging motion capture and 3D scanning for high-fidelity facial performance acquisition. SIGGRAPH, 2011
- [42] Sergios Theodoridis. 模式识别（第4版）. 李晶皎等译. 电子工业出版社, 2012
- [43] 李航著. 统计学习方法. 北京：清华大学出版社, 2012
- [44] 翁浩, 贾金原. 单张图片树木 L-system 的智能提取算法 [J]. 计算机科学与探索, 2013, 7 (2): 145~151
- [45] 项亮. 动态推荐系统关键技术研究（PhD thesis）. 中国科学院, 2011
- [46] The Future of 3D Printing Services and Manufacturing. CSC. <http://www.csc.com>
- [47] Ultimaker Wiki. http://wiki.ultimaker.com/Ultimaker_wiki
- [48] 3D 打印编年史——从 19 世纪末说起. 筑梦创造. <http://www.mongcz.com/archives/5904>
- [49] 德国制造：从“山寨”到“标杆”. 南京日报. 2013-1-14
- [50] 那些可以被 3D 打印的材料们. 方片 3. <http://www.fangpian3.com>, 2013
- [51] 金属 3D 打印技术小盘点. 方片 3. <http://www.fangpian3.com>, 2013
- [52] Shapeways. <http://www.shapeways.com/>
- [53] Thingiverse. <http://www.thingiverse.com/>
- [54] Autodesk. <http://www.autodesk.com/>
- [55] 3d printer joysmaker Q&A. 乐享 3D. <http://www.3djoy.cn>

- [56] 众筹网站领军者 Kickstarter. 快鲤鱼 . <http://www.kuailiyu.com>
- [57] Quirky : 创意加工厂 . 创业邦 . <http://www.cyzone.cn>
- [58] 四轴飞行 diy 全套入门教程 . 电子工程师学习交流园地 . <http://www.eeboard.com>
- [59] 三城记 : 中国创客地图 . 商业价值, 2013-7-3
- [60] 3D 打印挑战法律秩序 . 检察日报, 2013-8-16
- [61] 你注入灵感, 它铸就现实——3D 打印专题报告 . 华泰证券, 2013-3-12
- [62] 樊彬 . 局部图像特征描述概述 . 视觉计算研究论坛 . <http://www.sigvc.org/bbs/>
- [63] 颜永年, 齐海波, 林峰等 . 三维金属零件的电子束选区熔化成形 . 机械工程学报, 2007, 43 (6) : 87~92
- [64] 史玉升, 鲁中良, 章文献等 . 选择性激光熔化快速成形技术与装备 . 中国表面工程, 2006, 19 : 150~158
- [65] 编码无悔 -Intent & Focused. <http://www.codelast.com>
- [66] 袁亚湘, 孙文瑜著 . 最优化理论与方法 . 科学出版社, 1997
- [67] 陈宝林编著 . 最优化理论与算法 (第 2 版) . 清华大学出版社, 2005
- [68] 孟岩 . 博客频道 : <http://blog.csdn.net/myan/>
- [69] Thabo Beeler, Bernd Bickel, P. Beardsley, B. Sumner. High-Quality Single-Shot Capture of Facial Geometry. SIGGRAPH, 2010
- [70] R. Sumner, J. Popović. Deformation Transfer for Triangle Meshes. SIGGRAPH, 2004
- [71] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11): 1330-1334 (2000)
- [72] Jose I. Echevarria, Derek Bradley, Diego Gutierrez, Thabo Beeler. Capturing and Stylizing Hair for 3D Fabrication. SIGGRAPH, 2014
- [73] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, Jovan Popovic. Face Transfer with Multilinear Models. SIGGRAPH, 2005

后 记



《周易·系辞上》有云：“书不尽言，言不尽意”。敲完本书正文章节的最后一个字符，蓦然发觉已别十月之秋。自己的想法和思路虽然很多已写成了文字，但离真正诠释还相去甚远，受本人之才力和精力所限，已砉然笔落。

要用最简单的文字，把 3D 打印和 3D 智能数字化这些专业领域的理论和方法讲清楚，实在是个非常巨大的挑战。幸好，自然界的本质可以利用简单的、线性的方法去认识，所有复杂的、非线性的理论都可用简单的、线性的模型去逐段拟合或迭代逼近。正如牛顿在《自然哲学的数学原理》一书指出：“Natura enim simplex est, et rerum causis superfluis non luxuriat.（大自然偏爱简单，而不喜欢多余的浮夸）”。爱因斯坦也说过：“力图把表面上极为复杂的自然现象归结为几个简单的基本概念和关系，这就是整个自然哲学的基本原理”。

正因为如此，笔者不想让这本书成为一本晦涩难懂的“非读物”，而是希望能够让零基础的人也可以循序渐进地、从第 1 页一口气读到最后一页，直到掌握那些看起来复杂深奥的理论方法。实际上，无论 3D 打印，还是 3D 智能数字化（视觉计算、模式识别、机器学习）都是非常有趣、非常实用的学科。做科研和做产品，也理应从感兴趣开始！

笔者相信，在这个“英雄出少年”的时代，借助“取之不尽、用之不竭”的网络知识共享和协作，在本书的读者当中，将来很有可能会涌现出一些能比肩或超越比尔·盖茨和乔布斯的青年才俊。他们虽然现在是零基础，但身体里却充满着创造的活力和热情。因此，本书正是希望告诉年轻的创客们，别看 3D 打印和 3D 智能数字化的专业理论和方法似乎很高深，但其实所蕴含的数学工具和方法都很美，抛去公式化的抽象后，本质上都很简单与和谐。

随着 3D 打印产业链的不断升级和完善，“个人智造”、“家庭智造”、“网络社区智造”将登上历史舞台。创新思想的潮流一旦被开闸放出，打上自由精神烙印的个性化定制产品将会迅速充斥整个社会的各个角落。未来，将是一个缤纷五彩，甚至满目纷杂的后工业化时代。如果，你相信这个世界存在“道”，那么，“道”从未改变过，只是你的“心”可能会因环境而变了。因此，理清各种关系和渊源之后，对世界本质的探究依然可归结为简单的线索。正如第 1 章提到的，将来我们开发一个新产品时，不了解某一相关领域的知识，没关系，我们可以去创客社区（Maker Community）或智能云网（ICN, Intelligent Cloud Network）购买现成的标准技术组件（称之为创件 Makeware）。各种复杂的内部机制都被封装，我们只需调用清晰定义的接口即可，无须弄明白创件里面的每一个学科知识细节，这样才“简单而有效”。

未来，3D 打印及 3D 智能数字化技术将会变得越来越复杂，学习和掌握的难度也会变得越来越大，但是，请不要慌乱！因为人类对世界的认知符合奥卡姆剃刀定律（Occam's Razor），也即“简单有效原理”。实际上，无论是在科学领域、经济领域还是整个社会领域，只有简单有效的原理或操作模式才最容易为人们所广泛接受，并成为前进发展的基石。因此，把书本中的基本原理和方法掌握好，无论面对多么复杂的技术和技巧，都可如庖丁解牛般胸有成竹、游刃有余。

《通书·文辞》有云：“文以载道”。本书力图使用通俗易懂的文字及符号来阐述 3D 打印及 3D 智能数字化之道，以便让你有兴趣把整本书都读完，并明白其中的方法原理，掌握最新的技术技巧，此乃笔者平生之一大幸事。

与本书笔者交流可访问：

- 个人网址：<http://www.sigvc.org/why/>
- 个人微博：<http://weibo.com/huaiyuwu>
- 交流论坛：<http://www.sigvc.org/bbs/>

作者简介

吴怀宇，江西丰城人，中国科学院副研究员、中国 3D 科技创新产业联盟副理事长。任职于中国科学院自动化研究所，模式识别国家重点实验室（NLPR），中国 - 欧洲信息、自动化与应用数学联合实验室（LIAMA）。2008 年 7 月在中国科学院获“模式识别与智能系统”专业博士学位。2008 年 7 月至 2011 年 8 月在北京大学信息学院做博士后、讲师，并膺获“北京大学优秀博士后”称号。2011 年 8 月至今，在中国科学院自动化研究所模式识别国家重点实验室任助理研究员、副研究员。目前担任多个国际刊物的评审专家和国际程序委员会成员等学术任职，是美国电气与电子工程师学会（IEEE）、美国计算机协会（ACM）、IEEE Computer Society 会员，并担任过 ICCV/ CVPR/ACCV 国际程序委员会委员、程序主席秘书，以及北京市科学技术委员会项目评审专家、国家自然科学基金评审专家、国家科技计划高新领域评审专家。



主要研究领域包括 3D 智能数字化打印、计算机三维视觉、视觉形状感知分析与处理、计算机交互式图形学等。主持和参与国家自然科学基金（2 项，其中因取得突出研究进展获首批国家青年科学基金 - 面上项目连续资助项目）、国家高技术研究计划 863 项目（4 项）、中国博士后科学基金（一等）、北京市自然科学基金等多项国家重大科研课题。在计算机视觉 / 计算机图形学领域的 IEEE Transactions on Visualization and Computer Graphics、IEEE Transactions on IP、IEEE Transactions on CSVT、IEEE Transactions on ITS、ICCV、CVPR 等国际顶级期刊 / 会议上发表学术论文 30 余篇，相关技术申请国际 / 国家发明专利 5 项。研究成果应用到国产 3D 影视动漫制作当中，如国产三维动画电影《麋鹿王》中的三维形状渐变，该动画片获得第 13 届中国电影华表奖优秀动画片奖、第 27 届中国电影金鸡奖最佳美术片提名奖。

作为我国 3D 智能数字化打印领域的前沿人物，受《中国科学报》、《光明日报》、《中国自动化学会通讯》邀请撰写长篇技术评论并连载在 2013 年的最新版面上，标题分别为：《3D 数字化与 3D 打印：用“虚拟”再造“现实”》、《3D 数字化与 3D 打印：转向“中国智造”的产业机遇》、《智能数字化与 3D 打印：“中国智造”推动“全球第三次工业革命”》、《3D 打印：智能数字化》。先后主持和参与高精度三维数码照相与智能立体打印系统、真实感三维人脸建模和编辑关键技术、基于随机回归森林与多源数据融合的高精度三维动态形状获取、流形调和分析的三维形状匹配与检索、立体视觉方法的三维运动捕捉系统研究及其应用、面向复杂非规则多运动对象的大规模全景动态光场采集与再现系统、虚实融合协同工作的集成环境和关键技术的科研工作。